

Using the U.S. Climate Reference Network to Identify Biases in Near- and Subsurface Meteorological Fields in the High-Resolution Rapid Refresh (HRRR) Weather Prediction Model

TEMPLE R. LEE^a, RONALD D. LEEPER,^{b,c,d} TIM WILSON,^{a,c} HOWARD J. DIAMOND,^a TILDEN P. MEYERS,^a AND DAVID D. TURNER^f

^a NOAA/Air Resources Laboratory, Oak Ridge, Tennessee

^b North Carolina Institute for Climate Studies, Asheville, North Carolina

^c NOAA/National Centers for Environmental Information, Asheville, North Carolina

^d Center for Weather and Climate, Asheville, North Carolina

^e Oak Ridge Associated Universities, Oak Ridge, Tennessee

^f NOAA/Global Systems Laboratory, Boulder, Colorado

(Manuscript received 9 December 2022, in final form 23 March 2023, accepted 27 March 2023)

ABSTRACT: The ability of high-resolution mesoscale models to simulate near-surface and subsurface meteorological processes is critical for representing land–atmosphere feedback processes. The High-Resolution Rapid Refresh (HRRR) model is a 3-km numerical weather prediction model that has been used operationally since 2014. In this study, we evaluated the HRRR over the contiguous United States from 1 January 2021 to 31 December 2021. We compared the 1-, 3-, 6-, 12-, 18-, 24-, 30-, and 48-h forecasts against observations of air and surface temperature, shortwave radiation, and soil temperature and moisture from the 114 stations of the U.S. Climate Reference Network (USCRN) and evaluated the HRRR's performance for different geographic regions and land cover types. We found that the HRRR well simulated air and surface temperatures, but underestimated soil temperatures when temperatures were subfreezing. The HRRR had the largest overestimates in shortwave radiation under cloudy skies, and there was a positive relationship between the shortwave radiation mean bias error (MBE) and air temperature MBE that was stronger in summer than winter. Additionally, the HRRR underestimated soil moisture when the values exceeded about $0.2 \text{ m}^3 \text{ m}^{-3}$, but overestimated soil moisture when measurements were below this value. Consequently, the HRRR exhibited a positive soil moisture MBE over the drier areas of the western United States and a negative MBE over the eastern United States. Although caution is needed when applying conclusions regarding HRRR's biases to locations with subgrid-scale land cover variations, general knowledge of HRRR's biases will help guide improvements to land surface models used in high-resolution weather forecasting models.


SIGNIFICANCE STATEMENT: Weather forecasters rely upon output from many different models. However, the models' ability to represent processes happening near the land surface over short time scales is critical for producing accurate weather forecasts. In this study, we evaluated the High-Resolution Rapid Refresh (HRRR) model using observations from the U.S. Climate Reference Network, which currently includes 114 reference climate observing stations in the contiguous United States. These stations provide highly accurate measurements of air temperature, precipitation, soil temperature, and soil moisture. Our findings helped illustrate conditions when the HRRR performs well, but also conditions in which the HRRR can be improved, which we expect will motivate ongoing improvements to the HRRR and other weather forecasting models.

KEYWORDS: Atmosphere-land interaction; Surface observations; Model evaluation/performance

1. Introduction

The High-Resolution Rapid Refresh (HRRR) model is a deterministic convection allowing model with a 3-km grid spacing over the continental United States that has been used to support operational weather forecasting since 2014 (e.g., Benjamin et al. 2016; Dowell et al. 2022; James et al. 2022). An accurate representation of near- and subsurface

meteorological processes, in addition to feedback processes between the land surface and overlying atmosphere, is critical for the HRRR and its successors, e.g., the Rapid Refresh Forecast System [RRFS, which is scheduled to replace the HRRR in 2024 (Dowell et al. 2022)], to produce reliable weather forecasts. Doing so requires that the model is able to properly represent land–atmosphere feedback processes. Previous studies have reported that land cover variations affect the surface energy balance which in turn influences the evolution of the planetary boundary layer (PBL), development of clouds and precipitation, and surface radiative forcing (e.g., Oke 1987; Stull 1988; Segal and Arritt 1992; Pielke 2001; Ek and Holtslag 2004; Leeper et al. 2009; Dirmeyer et al. 2012; Wulfmeyer et al. 2018; Santanello et al. 2019; Min et al. 2021;

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Temple R. Lee, temple.lee@noaa.gov

DOI: 10.1175/WAF-D-22-0213.1

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Tian et al. 2022). The interactions among these processes are oftentimes highly complex and nonlinear, yet the ability of a model to properly simulate them is critical for improving forecasts.

Only a few studies have investigated the performance of near-surface meteorological fields within the HRRR. Lee et al. (2019) found that version 2 of the HRRR had a positive 2-m temperature bias of around 2°C during the warm season, but a bias less than 1°C during the cool season compared to observations from micrometeorological towers installed in northern Alabama from 2016 to 2017. Lee et al. (2019) also reported that, despite these relatively small biases in the near-surface meteorological fields, there were large differences between the modeled and observed surface sensible, latent, and ground heat fluxes. Min et al. (2021) evaluated version 3 of the HRRR using observations from the New York State Mesonet and found that the HRRR had a dry and warm bias during the warm season, which they defined as 1 June–31 August, and that the HRRR had a significant positive bias in solar radiation under overcast sky conditions. Fovell and Gallagher (2020) also evaluated version 3 of the HRRR over the contiguous United States using observations from rawinsondes and from the Automated Surface Observing System (ASOS). Fovell and Gallagher (2020) reported good agreement between the HRRR and observed near-surface temperature and wind speeds at the earlier forecast periods, but larger, positive wind speed biases in the PBL in the 24-h HRRR forecasts.

In addition to the above studies evaluating the performance of the HRRR model itself, other studies have evaluated the Rapid Update Cycle (RUC) land surface model (LSM), which is the LSM used in the HRRR (e.g., Benjamin et al. 2016; Smirnova et al. 2016), by conducting simulations using the Weather Research and Forecasting (WRF) Model (e.g., Lee et al. 2016; Kantha Rao and Rakesh 2019; Alexander et al. 2022). Alexander et al. (2022) used the WRF Model to evaluate the performance of different LSMs and PBL parameterization schemes over California's Central Valley. The researchers found a stronger sensitivity to the choice of LSM than PBL scheme, but they found that the Pleim–Xiu LSM (Xiu and Pleim 2001; Pleim and Xiu 2003) better simulated near-surface turbulent fluxes and humidity. Lee et al. (2016) similarly showed strong sensitivity to the choice of LSM and found that the Noah LSM (Ek et al. 2003) compared better against observations than the other LSMs they evaluated (i.e., the RUC LSM and Pleim–Xiu LSM).

Unfortunately, there are no known studies that have evaluated how well the HRRR simulates soil moisture and temperature variables across the contiguous United States. Soil moisture plays a vital role in the partitioning of radiation into sensible and latent heat, which is a critical process to represent accurately in LSMs (e.g., Santanello et al. 2019; Min et al. 2021). Knowledge of how well the HRRR simulates soil moisture can help to inform improvements to the model's representation of exchanges of heat and moisture between the land and atmosphere. Furthermore, characterizing the HRRR's performance for both below- and above-ground meteorological variables, as well as evaluating the HRRR's performance over different regions of the United States and for different land cover types, is important to guide

improvements to the HRRR, and ultimately the RRFs, as well as to the LSMs used therein. Improvements to these modeling systems will reduce uncertainties in short-term weather forecasts and water supply forecasts.

2. Datasets

a. USCRN

In this study, we evaluated the HRRR from 1 January 2021 to 31 December 2021 using data from the stations comprising the U.S. Climate Reference Network (USCRN; Diamond et al. 2013) within the contiguous United States. The USCRN was commissioned in 2004, and installation of stations across the contiguous United States was completed in 2008 (Bell et al. 2013). The aim of the USCRN is to provide long-term, accurate observations of air temperature, precipitation, soil moisture and temperature from sites undisturbed by changes in land cover (e.g., urbanization). Since the network's inception, data from the USCRN have been used for drought monitoring (e.g., Leeper et al. 2021), as a component of a study of the Great American eclipse in 2017 (e.g., Lee et al. 2018), the development of heat exposure products (Rennie et al. 2021), validation of satellite products (e.g., Otkin et al. 2005; Krishnan et al. 2015; Gallo and Krishnan 2022), evaluations of meteorological observations from other networks (e.g., Hubbard et al. 2004; Sun et al. 2005; Leeper et al. 2015), and model verification studies (e.g., Leeper et al. 2017; Buban et al. 2020).

Each USCRN station is installed in an open area and consists of air temperature, relative humidity, precipitation, and wind speed that are sampled 1.5 m above ground level (AGL), in addition to surface (or skin) temperature and incoming shortwave radiation (e.g., Bell et al. 2013; Diamond et al. 2013). Air temperature is measured using three independent Thermometrics platinum resistance thermometers installed within a fan-aspirated radiation shield, and surface temperature is obtained using an Apogee Instruments infrared temperature sensor. Previous studies have shown that surface temperatures derived from USCRN compare well against surface temperatures derived from other platforms such as pyrometers on board aircraft (e.g., Krishnan et al. 2015). Precipitation is measured at each USCRN station using a Geonor T-200B precipitation gauge. To optimize the catch efficiency of solid precipitation, a small double-fence inter-comparison reference shield (SDFIR) surrounds each gauge (e.g., Rasmussen et al. 2012). Solar radiation observations are made using a Kipp and Zonen SP Lite pyranometer. Based on repeated calibrations for each of these sensors and the age of the sensors, the uncertainty of the solar radiation observations is on the order of 10%. Thus, to gain additional confidence in the USCRN radiation measurements, we compared these measurements against collocated radiation measurements from the SURFRAD station in Champaign, Illinois (Augustine et al. 2000). Over the 1-yr period of interest, there was excellent agreement between the hourly mean USCRN radiation observations and hourly mean incoming shortwave radiation observations from SURFRAD; R^2 was 1.00, and the mean bias was $-3.06 \pm 11.1 \text{ W m}^{-2}$.

Soil moisture and temperature measurements were added to the USCRN starting in 2009, and installation was completed in 2011 (e.g., [Diamond et al. 2013](#)). Soil moisture and temperature are measured in triplicate profiles at standard depths of 5, 10, 20, 50, and 100 cm below ground using the HydraProbe II soil sensors produced by Stevens Water Monitoring Systems, Inc. as well as the TDR-315 sensors from Acclima, Inc. (e.g., [Bell et al. 2013](#); [Wilson et al. 2020](#)). Recent work has shown that measurements from the HydraProbes compare well on daily time scales with measurements from Acclima's 1-GHz time domain reflectometry (TDR) sensor, but we acknowledge that the TDR sensor was found to provide more representative soil moisture measurements than the HydraProbe in soils with high clay content ([Wilson et al. 2020](#)).

Monthly, daily, hourly, and subhourly data from the USCRN are publicly available (<https://www.ncei.noaa.gov/access/crn/qcdatasets.html>). In this study, we used hourly data from the USCRN to evaluate the HRRR model performance. Although the datasets from USCRN are provided as a continuous time series, there are occasional periods of missing data resulting from unavoidable station outages. In the present study, we ignored observations for a given variable at a given station if more than 25% of the record was missing. As shown in [Table 1](#), 99.1% and 97.4% of the stations have at least 75% of the data available for air temperature and surface temperature, respectively, and 97.4% of the stations have at least 75% of the precipitation and incoming shortwave radiation available. The percent data completion is lower for soil temperature and soil moisture. For example, 84.5% and 69.8% of the USCRN stations have at least 75% of the soil temperature and soil moisture data, respectively, available for the 5-cm depth which is largely due to e.g., frozen soils during the cold season that limit the availability of soil moisture observations.

b. HRRR

The HRRR has 1799×1059 grid points due to its 3-km grid spacing across the continental United States. The land use/land cover data has a 15-s (~ 0.5 km) resolution, and the leaf area index data are estimated from the Moderate Resolution Imaging Spectroradiometer (MODIS) (e.g., [Min et al. 2021](#)). The HRRR itself uses the RUC LSM, which has nine vertical levels of soil moisture and soil temperature

[0 (i.e., the surface or skin value), 1, 4, 10, 30, 60, 100, 160, and 300 cm below ground level]. The RUC computes fluxes using an implicit scheme (e.g., [Smirnova et al. 1997](#); [He et al. 2021](#)) and uses Richard's equation to represent water transport through the soil ([Richards 1931](#); [Smirnova et al. 1997](#); [He et al. 2021](#)), while heat is transferred one dimensionally through the soil (e.g., [Smirnova et al. 1997](#); [He et al. 2021](#)). Recent updates to the RUC LSM include an improved treatment of snow through the implementation of a two-layer snow model and also better representations of melting snow ([Smirnova et al. 2016](#)). Updates to version 4 of the HRRR include improvements to the HRRR's data assimilation systems, specifically through the implementation of a moderately coupled land-snow-atmosphere assimilation method ([Benjamin et al. 2022](#)). In addition to using the RUC LSM, the HRRR uses the Mellor-Yamada-Nakanishi-Niino PBL scheme ([Nakanishi and Niino 2004, 2009](#); [Olson et al. 2021](#)), which has been found to have a better representation of subgrid scale clouds than previous HRRR versions (e.g., [Olson et al. 2019](#); [Dowell et al. 2022](#)). Additionally, the HRRR uses the Rapid Radiative Transfer Model Global (e.g., [Iacono et al. 2008](#)) and Thompson scheme (e.g., [Thompson and Eidhammer 2014](#)) for radiation and microphysics, respectively. We refer readers to e.g., [Dowell et al. \(2022\)](#) for more details about the HRRR's configuration.

We evaluated the latest operational version of the HRRR (i.e., version 4) for eight different forecast periods: the 1-, 3-, 6-, 12-, 18-, 24-, 30-, and 48-h forecasts. These forecasts were obtained from the Amazon web services' HRRR data archive (<https://registry.opendata.aws/noaa-hrrr-pds/>) using a data extraction protocol by [Blaylock et al. \(2017\)](#). Hourly HRRR output is available for the forecasts through 18 h, and the HRRR output is available every six hours for forecast periods beyond the 18-h forecast. Thus, in a given day, a forecast out to 18 h is provided every hour; however, at 0000, 0600, 1200, and 1800 UTC the model outputs hourly forecasts out to 48 h. Consequently, during the 1-yr study period, for the 1–18-h forecasts, there are 8760 different modeling runs for each of these forecast periods; for the 24-, 30-, and 48-h forecasts, there are 2190 different modeling runs for each of these forecast periods.

To evaluate the performance of the model within the soil, we used the HRRR output from 4 and 10 cm for comparison against the USCRN soil moisture and soil temperature observations, which closely align with the USCRN's 5- and 10-cm measurements. Because sensors were not installed at 100 cm at all USCRN stations, we focus only on the 5- and 10-cm depths in this study.

3. Methods

For all variables except for precipitation, we computed the daily means from the hourly values at each USCRN station and compared these values against the modeled values obtained from the HRRR grid cell containing that station. We acknowledge, though, that we did not treat stations that were located very near the edge or corner of the HRRR grid cell in

TABLE 1. Percent of stations with at least a 75% data completion for a given variable.

Variable	Percent
Air temperature (T_a)	99.1
Skin temperature (T_s)	97.4
Incoming shortwave radiation (SW_d)	98.3
Precipitation (P)	97.4
5-cm soil temperature (ST_{05})	84.5
10-cm soil temperature (ST_{10})	87.1
5-cm soil moisture (SM_{05})	69.8
10-cm soil moisture (SM_{10})	75.0

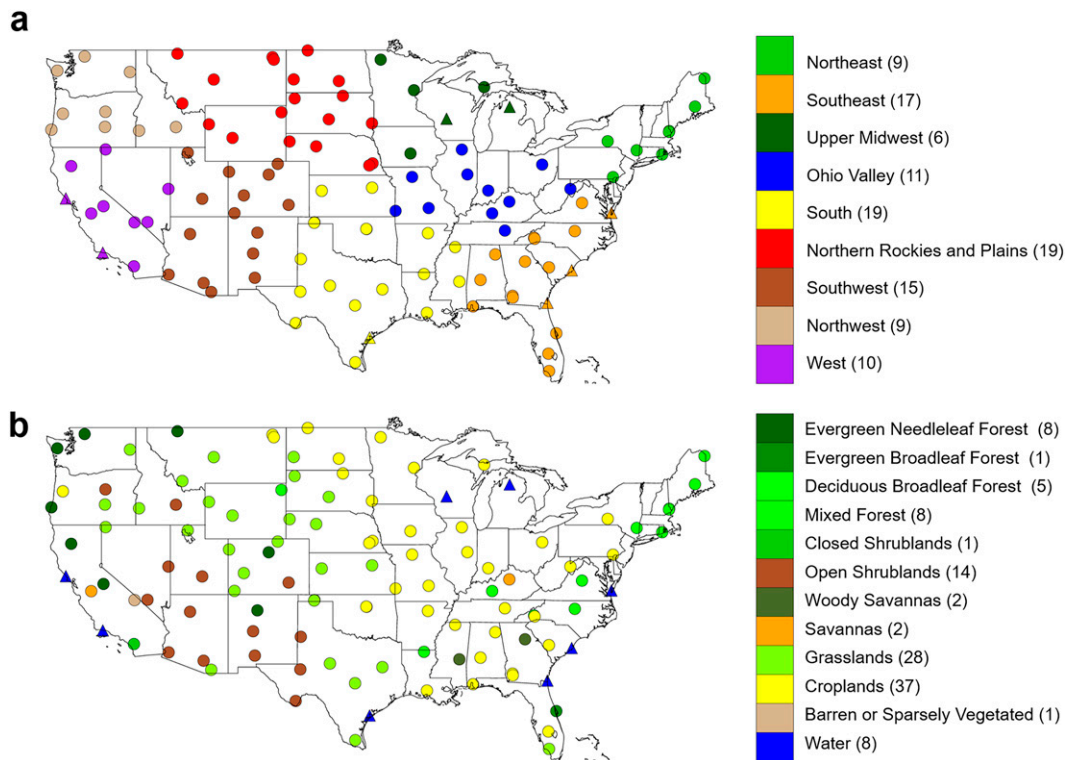


FIG. 1. USCRN stations classified by (a) NCEI climate region and (b) the station's land cover type based on the HRRR grid cell. Stations used in the analyses are denoted by a filled circle, and the stations that are omitted from all analyses because of the water land cover classification are denoted by a filled triangle. The number of stations for each category is shown in parentheses in the legend.

which the USCRN station was located any differently from stations located nearer the center of the HRRR grid cell. Because we averaged across multiple stations and seasons, we expect any affects here to be minimal. We also note the known caveats associated with comparing a point observation from each USCRN station with a 3-km grid cell. For example, in flat, homogeneous areas, the difference in scale between a point observation and model grid cell is less important than for stations located in complex topography and/or cases where there exists subgrid-scale land cover variations. We acknowledge that the uncertainty of comparing the point measurements from the USCRN stations will be larger in the latter instance, although we do not explicitly quantify this uncertainty in the present study.

Comparing model-derived precipitation with observed precipitation presents its own unique challenges. Because slight misplacements in precipitation can lead to the so-called double penalty problem (e.g., Gilleland et al. 2009; Ikeda et al. 2013), we first computed the sum of the hourly precipitation totals for each day and determined the maximum modeled value over an $18 \text{ km} \times 18 \text{ km}$ area (i.e., 36 HRRR grid cells) surrounding the station's grid cell. We then computed the daily sum of the maximum modeled values and used this daily sum for comparison against the observed daily precipitation total at each USCRN station.

We then evaluated the performance of the HRRR for all variables by computing the mean bias error (MBE), which we define as the output from the model for a given variable minus the observed value; the slope of the best fit line between the modeled and observed values (m_b); the coefficient of regression (R^2); and the root-mean-square error (RMSE). To assess the performance of the HRRR over different regions of the United States, we classified the USCRN stations into the nine climate regions defined by the National Center for Environmental Information (NCEI): Northeast, Southeast, Upper Midwest, Ohio Valley, South, Northern Rockies and Plains (i.e., West North Central), Southwest, Northwest, and West (Fig. 1a).

We also evaluated the performance of the HRRR across different land cover types by using the land cover type derived from the HRRR grid cell containing each USCRN station. The HRRR uses 21 land cover categories which are obtained from MODIS; of the 21 land cover categories, 12 are represented by at least one USCRN station. However, for the purposes of this study, we classified stations with the HRRR land cover closed shrublands and open shrublands as shrublands, and we reclassified stations with the HRRR land cover woody savanna and savanna land cover types as savanna. Additionally, we classified stations with the HRRR land cover evergreen needleleaf forest, evergreen broadleaf forest, deciduous broadleaf forest, and mixed forest as simply forest. As shown in Fig. 1b, however, the majority of USCRN stations are

classified as either croplands or grasslands; these land cover types account for the land cover at 32% and 24%, respectively, of the USCRN stations.

Eight of the USCRN stations are classified as the water land cover type in the RUC LSM due to the stations' proximity to the Atlantic Ocean, Pacific Ocean, Great Lakes, or in the case for the USCRN station at Necedah, Wisconsin, due to several lakes near the station (cf. Fig. 1b). Since the pixels with the water land cover classification are treated differently in the RUC LSM as compared with the other land cover types, we removed these eight USCRN stations with the water land cover classification from all subsequent analyses. Furthermore, when evaluating the HRRR performance over different land cover types, we omitted the barren or sparsely vegetated land cover classification because there was only one USCRN station with this particular land cover type, although we did include this station in all other analyses.

As previous studies have reported a warm, dry bias in the HRRR (e.g., Benjamin et al. 2016; Lee et al. 2019; Min et al. 2021), in this study we evaluated the role of radiation on the model MBE by computing a clearness index C_{index} for each day in 2021 using the incoming shortwave radiation (SW_d) observations obtained from the USCRN stations. The C_{index} is calculated as follows:

$$C_{index} = \frac{\sum SW_o}{\sum SW_t} \quad (1)$$

In Eq. (1), $\sum SW_o$ is the total amount of incoming shortwave radiation measured at a USCRN site and summed for a given day; $\sum SW_t$ is the sum of the total theoretical maximum incoming solar radiation that could be received on that day at that particular location and is calculated using the procedure described by Whiteman and Allwine (1986). Following Whiteman and Allwine (1986), SW_t can be calculated for any location on Earth's surface and varies as a function of time of day, day of year, latitude, longitude, and the inclination, azimuth, and steepness of the slope (e.g., Whiteman and Allwine 1986; Whiteman et al. 1999). Because the immediate area surrounding each USCRN station is relatively flat, we used a value of 0° for the inclination, azimuth, and steepness. Prior to computing the C_{index} , we removed physically unrealistic values of SW_d (e.g., nighttime values of $SW_d > 0$), as well as instances when the C_{index} is unrealistically high (e.g., $> \sim 0.85$ following Lee et al. 2015). Across all USCRN sites in 2021, the mean and median C_{index} was 0.51 and 0.57, respectively (Fig. 2), and the mean C_{index} for the different NCEI regions ranged from 0.45 in the Northeast and Upper Midwest to 0.60 in the Southwest.

4. Results and discussion

a. HRRR performance as a function of variable and forecast period

1) INCOMING SHORTWAVE RADIATION

We found that, when averaged across all USCRN stations in the contiguous United States, the difference between the

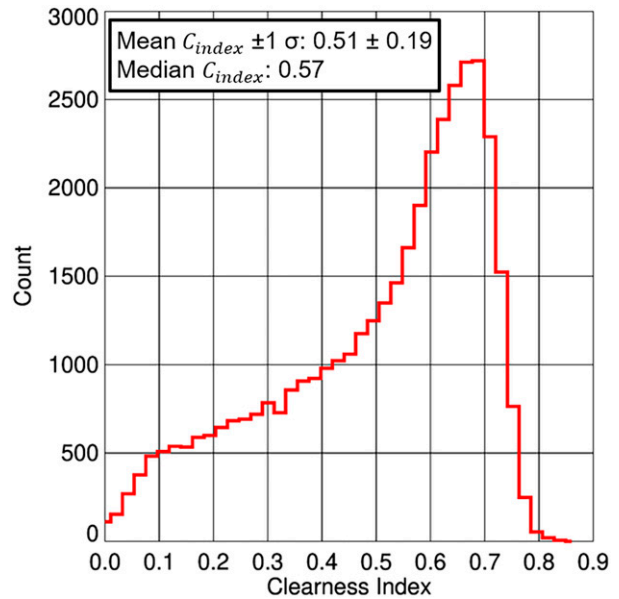


FIG. 2. The C_{index} histogram for all USCRN sites in 2021. The mean $C_{index} \pm 1$ standard deviation (i.e., σ) and median C_{index} are shown at the top left.

HRRR and observed values, averaged to daily time scales, varied seasonally for many of the variables evaluated in the present study. Mean SW_d had a positive MBE across the year and across all forecast hours that ranged from around 20 W m^{-2} in January and February to about 40 W m^{-2} between May and August (Fig. 3a). When averaged across the entire year and across all seasons, there was good agreement between the forecasted and observed SW_d , as R^2 and m_b were around 0.9 and 1, respectively, for the 1–18-h forecasts (Fig. 4). The R^2 and m_b decreased for the 24-, 30-, and 48-h forecasts but were still greater than 0.7 and 0.9, respectively. The MBE actually decreased as a function of forecast hour, but the standard deviation in the error increased as a function of forecast hour. When distinguishing by season, we found an R^2 and m_b of ~ 0.8 and ~ 0.9 , respectively, in January for the 1–18-h forecasts. Conversely, in the 24-, 30-, and 48-h forecasts, R^2 was 0.69, 0.53, and 0.59, respectively, and m_b was 0.83, 0.78, and 0.77, respectively (Fig. 5). Similarly, the larger MBEs occurred at the earlier forecast hours, but as when averaged over the entire year, the standard deviation in the error increased as a function of forecast hour. The same trend occurred in July, although July had lower values than January for R^2 and m_b , as these were around 0.7 and 0.8, respectively, for the 1–18-h forecasts (Fig. 6).

2) AIR TEMPERATURE

The difference between the modeled and observed air temperature also varied seasonally, although the difference lagged the SW_d MBE. Between January and May, as well as between October and December, the MBE generally ranged from about 0° to 0.5°C , although there were some instances during the latter forecast periods (i.e., the 30- and 48-h forecasts), when the daily

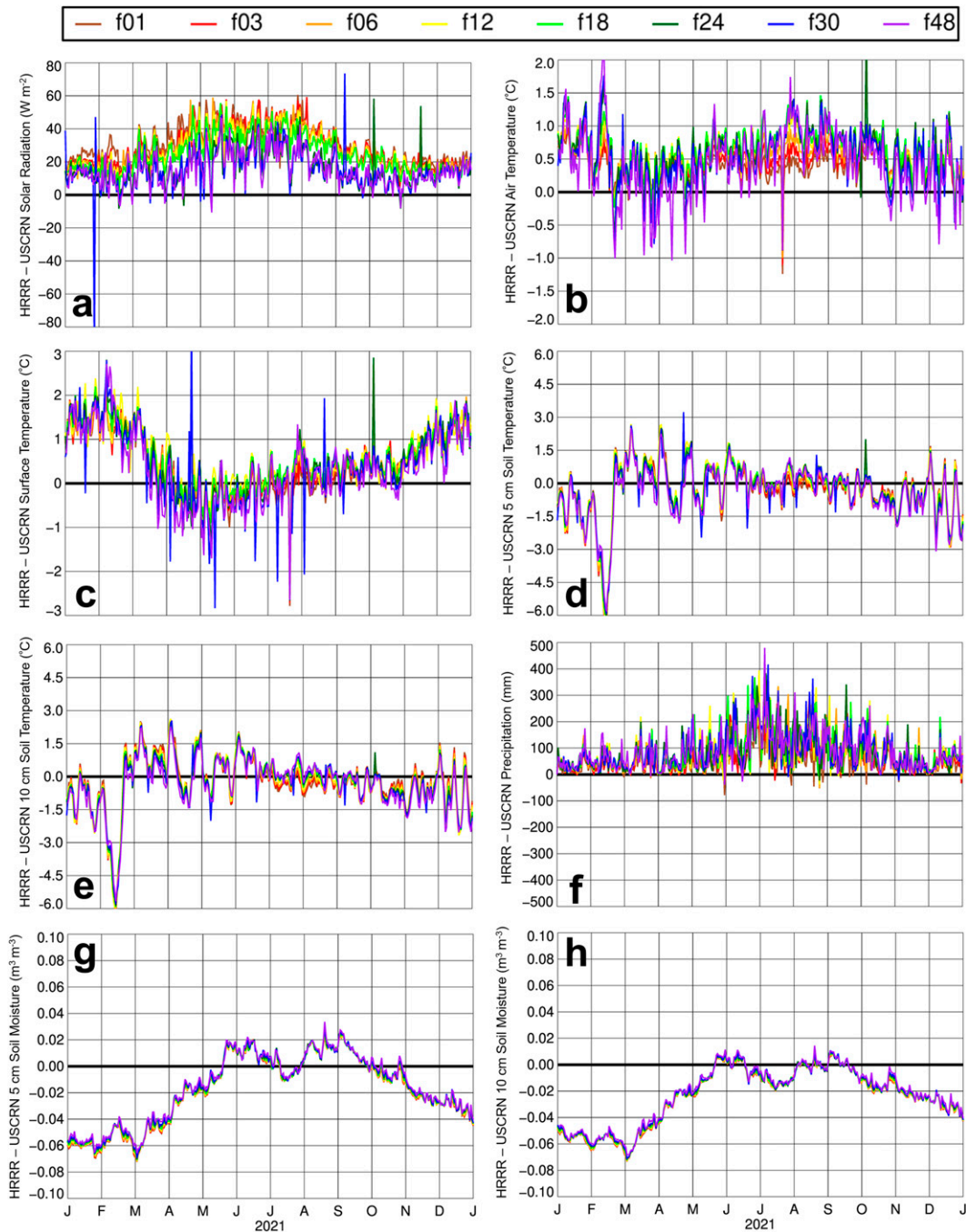


FIG. 3. Time series of mean HRRR minus mean USCRN observations as a function of time of year in 2021 for the 1- (brown line), 3- (red line), 6- (orange line), 12- (yellow line), 18- (green line), 24- (dark green line), 30- (blue line), and 48-h (purple line) HRRR forecasts for (a) SW_d , (b) T_a , (c) T_s , (d) ST_{05} , (e) ST_{10} , (f) P , (g) SM_{05} , and (h) SM_{10} . The values represent the mean daily differences over all of the USCRN stations, except for P , in which the daily totals are used both for the USCRN observations and HRRR model output prior to computing the differences between the model and observations.

MBE was low as -1.0°C . Between May and October, the MBE was between 0.5° and 1.0°C and was as large as 1.5°C in late July and August in the 30- and 48-h forecasts (Fig. 3b). For the year and for across all stations, the MBE was $0.35^\circ \pm 1.01^\circ\text{C}$ in the

1-h forecast but $0.37^\circ \pm 1.84^\circ\text{C}$ in the 48-h forecast and, as for SW_d , the standard deviation in the error increased as a function of forecast hour (Fig. 4). The R^2 and m_b were also remarkably high, as these were ≥ 0.97 and ~ 1.00 , respectively,

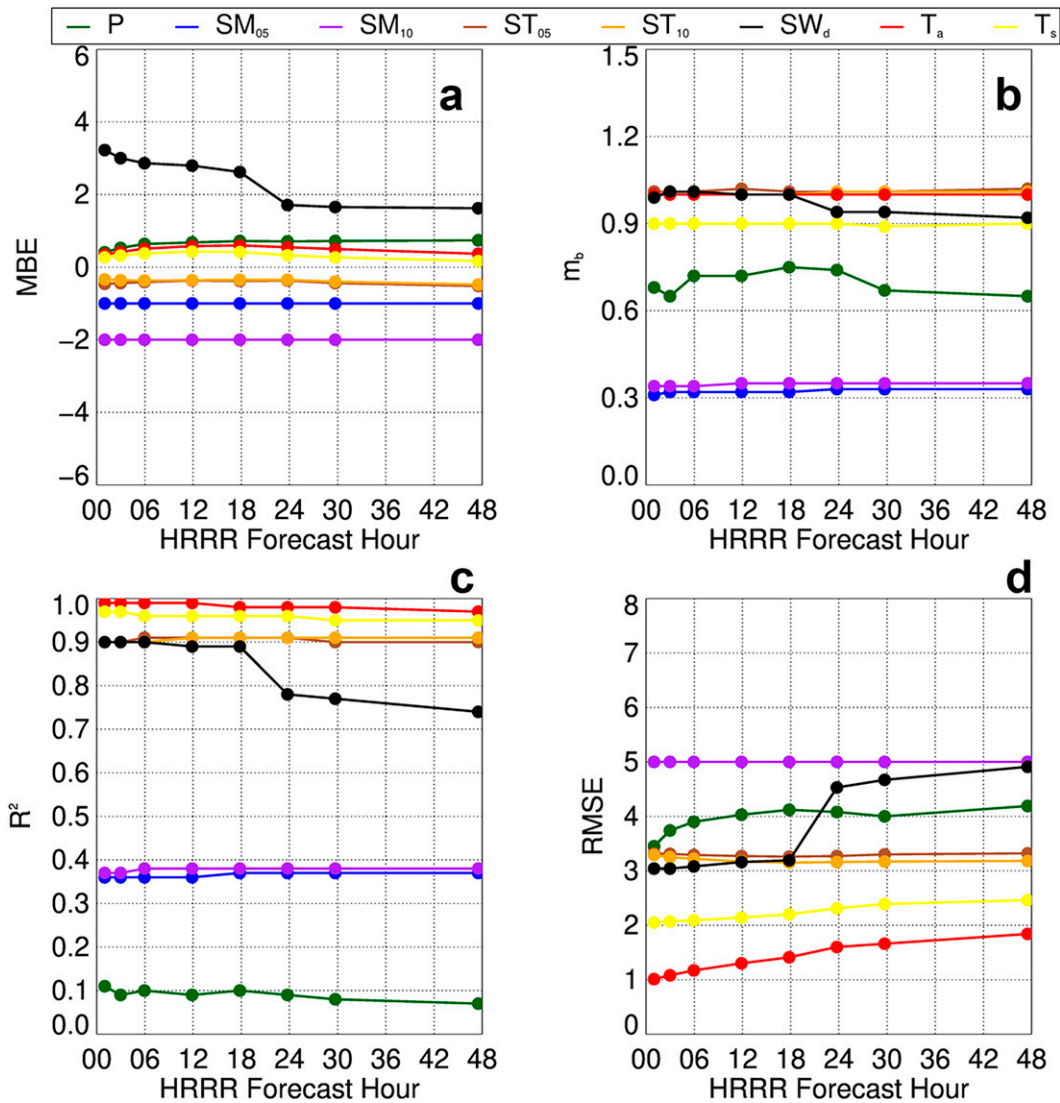


FIG. 4. (a) MBE, (b) m_b , (c) R^2 , and (d) RMSE for P (green line), SM_{05} (blue line), SM_{10} (purple line), ST_{05} (brown line), ST_{10} (orange line), SW_d (black line), T_a (red line), and T_s (yellow line) as a function of HRRR forecast hour. To show all lines on the same graphs, the MBE and RMSE for SM_{05} and SM_{10} have been multiplied by 100, and the MBE and RMSE for SW_d have been multiplied by 0.1. The units of MBE and RMSE for precipitation, soil moisture, temperature, and radiation are mm, $m^3 m^{-3}$, $^{\circ}C$, and $W m^{-2}$, respectively.

for all forecast hours. When differentiating by season, we found that these values were lower in July than in January. For example, R^2 was ≥ 0.92 for all forecast hours in January (Fig. 5) but decreased to 0.85 in the 48-h forecast in July (Fig. 6). Overall, the positive MBE that we found for SW_d and the ensuing positive MBE in air temperature has been reported in previous studies evaluating earlier versions of the HRRR and was attributed to the model’s inability to resolve subgrid-scale clouds (e.g., Benjamin et al. 2016; Lee et al. 2019; Wagner et al. 2019). We also note that, since the air temperature is a postprocessed field relying on relationships between the surface and first hybrid level above the surface rather than being explicitly calculated by the model, some of the errors in the HRRR’s air temperature may also be due to the postprocessing methodology used.

3) SURFACE AND SUBSURFACE TEMPERATURE

When assessing the HRRR’s performance for surface and subsurface temperature, we found lower values for m_b and R^2 than we found for air temperature. For skin temperature T_s , we found that m_b and R^2 were around 0.9 and 0.96, respectively, across all forecast hours (Fig. 4). Furthermore, these values were lower in July than in January. For example, m_b ranged from 0.87 to 0.81 between 1- and 48-h forecast, respectively, in January (Fig. 5), whereas in July m_b ranged from 0.77 to 0.72 between the 1- and 48-h forecast, respectively (Fig. 6). The MBE in the 5- and 10-cm soil temperatures did not show a clear seasonal pattern, although we note a period in which the MBE was $< -3^{\circ}C$ in February at both levels

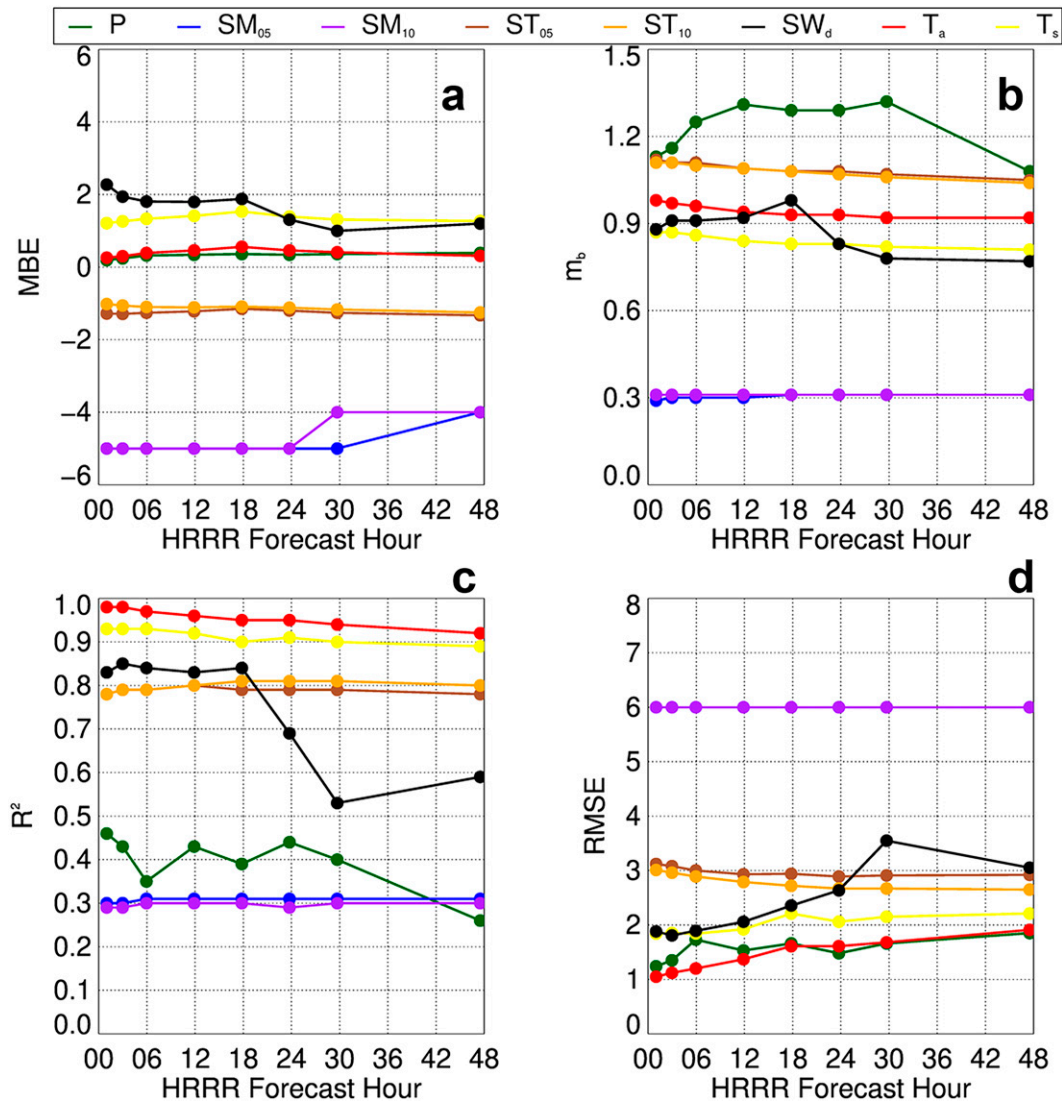


FIG. 5. As in Fig. 4, but for January 2021.

(Figs. 3d,e), which became more apparent when looking more closely at the relationship between the HRRR and USCRN observations, as shown in the example 1:1 plots in Fig. 7. Although we note a 1:1 relationship between the modeled and observed soil temperature for temperatures above 0°C, for temperatures below 0°C, the HRRR showed a much larger negative MBE in soil temperature, which was reported in previous work (Min et al. 2021) and has been attributed to the HRRR's inability to properly simulate freezing and thawing processes within the soil (e.g., Viterbo et al. 1999). The biases in soil temperature can also be attributed to the data assimilation approach used in the HRRR, as near-surface air temperatures are matched at the time of the model's initialization, and excess heat is moved into the soil (Benjamin et al. 2022). Another explanation for the negative MBE in soil temperature is that, in cases with snow, the HRRR permits partial snow cover, allowing for lower temperatures, whereas for

example a USCRN site may have full snow cover, providing insulation and resulting in higher observed soil temperatures than predicted by the HRRR. Overall, though, the agreement between the modeled and observed soil temperature did not vary significantly as a function of forecast hour; R^2 was ~ 0.9 for all forecast hours and for both levels, and m_b was ~ 1 for 5- and 10-cm soil temperature (Fig. 4). In January, m_b was ~ 1.1 for both depths and across all forecast hours (Fig. 5), whereas in July m_b was ~ 0.7 (Fig. 6).

4) PRECIPITATION

Errors in modeled precipitation increased as a function of forecast time, with an MBE of 0.41 mm in the 1-h forecast, but 0.74 mm in the 48-h forecast (Fig. 4) for the entire year. In January and July, the MBE ranged from 0.19 to 0.39 mm (Fig. 5) and from 0.76 to 1.36 mm (Fig. 6), respectively, between the 1- and 48-h HRRR forecasts. Model skill was also best for the 1-h

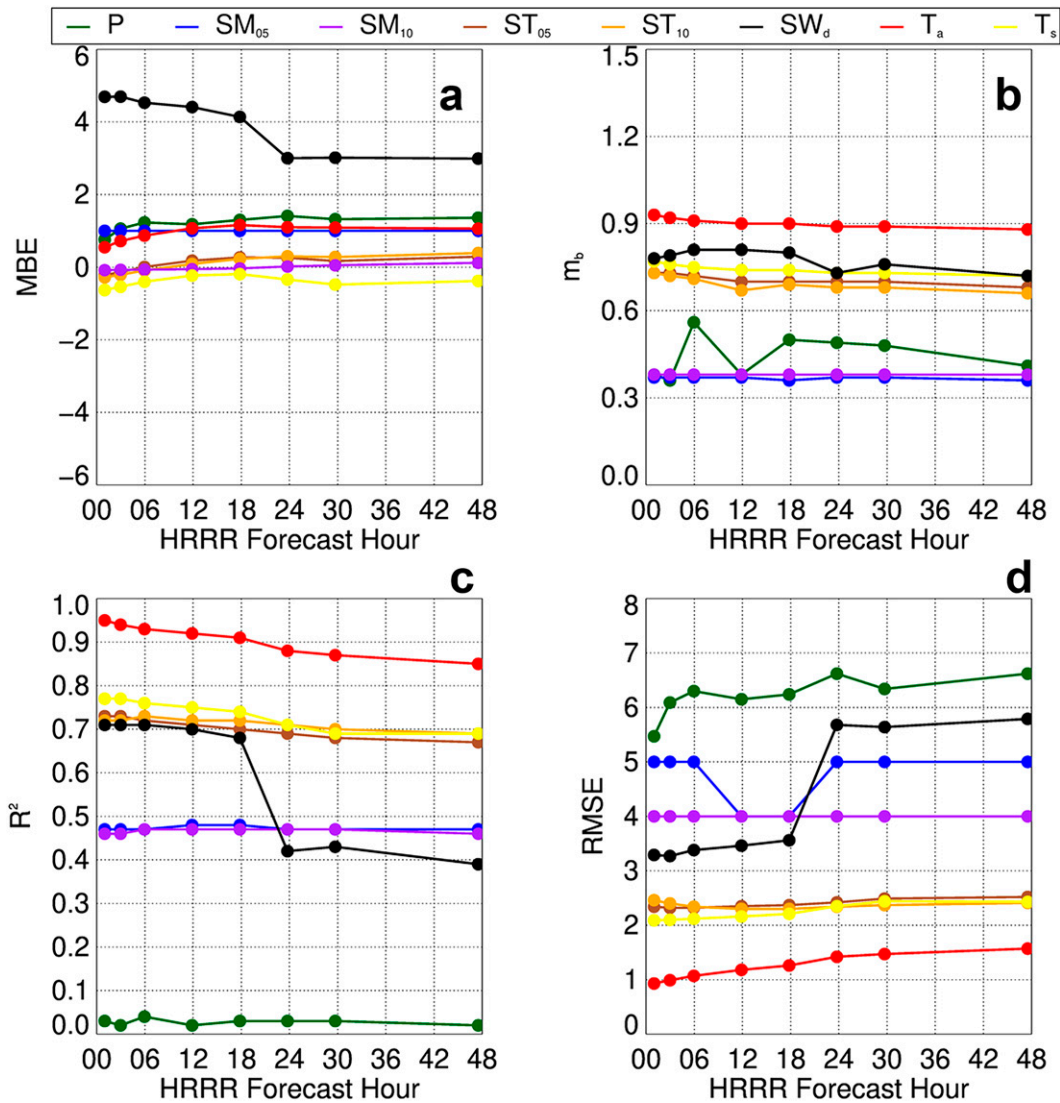


FIG. 6. As in Fig. 5, but for July 2021.

forecasts, as evident by the lowest RMSEs and generally largest values for R^2 .

5) SOIL MOISTURE

As was the case for soil temperature, when averaged across all USCRN stations in the contiguous United States, the MBE in soil moisture at 5 and 10 cm tracked each other very closely as a function of time of year. The largest MBE overall was in January and February, when the MBE was $0.05\text{--}0.06\text{ m}^3\text{ m}^{-3}$ (Figs. 3g,h). Between early May and November, the MBE across all forecast periods was $<\pm 0.02\text{ m}^3\text{ m}^{-3}$. The R^2 and m_b varied little as a function of forecast hour and were consistent for both depths, as R^2 and m_b were 0.36 and 0.34, respectively.

Overall, the HRRR generally overestimated soil moisture for dry soils (i.e., $<0.2\text{ m}^3\text{ m}^{-3}$) but underestimated soil moisture for more moist soils (i.e., those with observed soil

moisture $> 0.2\text{ m}^3\text{ m}^{-3}$), as illustrated in the 1:1 plots shown in Fig. 7. Previous studies using different models have found similar results, including Leeper et al. (2017) who evaluated the North American Regional Reanalysis (Mesinger et al. 2006) using USCRN data from 2011 to 2013. Other researchers reported similar errors in evaluations of the Noah LSM (e.g., Fan et al. 2011; Xia et al. 2015). One possible explanation for the errors is that the models do not account for the aggregate fraction (i.e., stones, etc.) for many of the soils in the western United States. For example, an aggregate fraction of 10% reduces the field capacity and soil moisture by 10%, which is similar to the MBEs we found.

b. Effects of clouds on HRRR performance

As noted earlier, when averaged across diurnal time scales, the HRRR generally overestimated SW_d for all seasons. We further investigated this behavior by distinguishing between

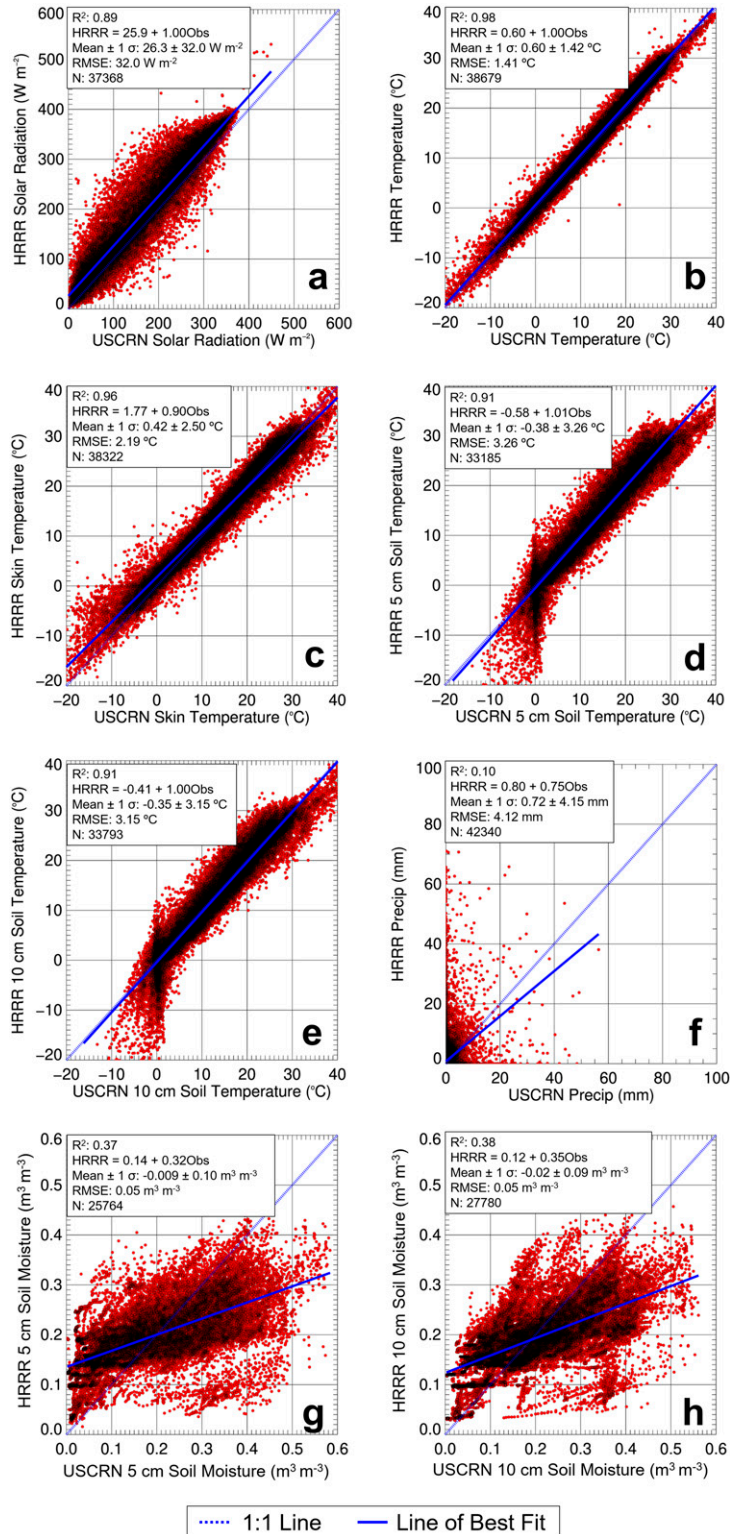


FIG. 7. Relationship between the HRRR 18-h forecast and USCRN observations of (a) SW_d, (b) T_a, (c) T_s, (d) ST₀₅, (e) ST₁₀, (f) P, (g) SM₀₅, and (h) SM₁₀. The solid blue line shows the line of best fit, and the dashed blue line shows the 1:1 line. The R², equation for the line of best fit, MBE ± 1 standard deviation (1σ) in the MBE, RMSE, and number of samples (N) are shown at the top left of each panel.

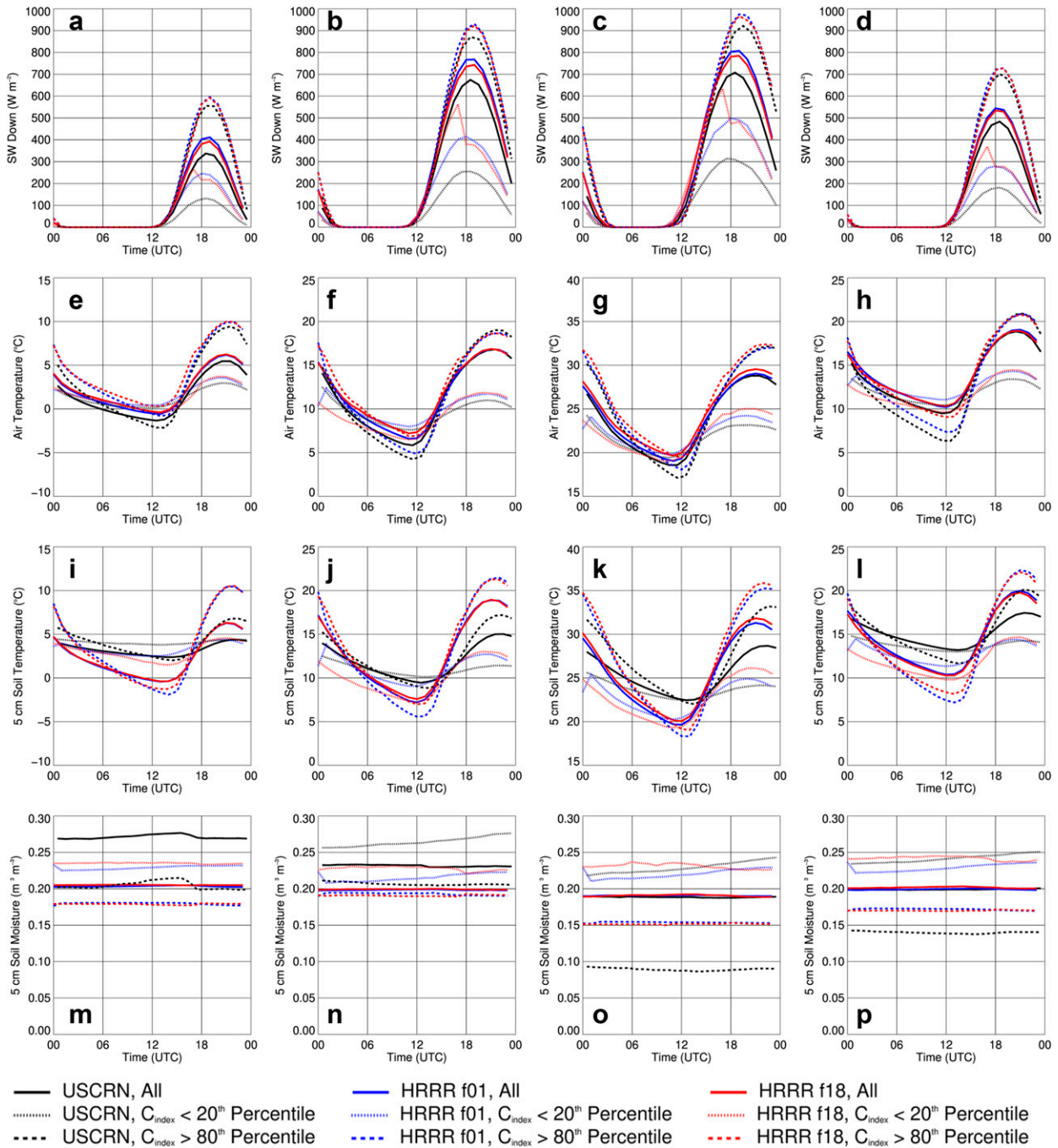


FIG. 8. Mean diurnal cycle of SW_d in (a) January, (b) April, (c) July, and (d) October for the USCRN (black line), the HRRR 1-h forecast (blue line), and the HRRR 18-h forecast (red line). (e)–(h), (i)–(l), (m)–(p) As in (a)–(d), but for T_a , ST_{05} , and SM_{05} , respectively. The solid lines represent the mean for all days; the dotted lines are only for days on which the C_{index} was below the 20th percentile; and the dashed lines are only for days on which the C_{index} was above the 80th percentile.

clear and cloudy days by computing C_{index} percentiles after removing days on which the C_{index} exceeded 0.85 (cf. section 3). We defined clear days as those on which the C_{index} exceeded the 80th percentile across all stations (i.e., $C_{index} > 0.684$), and we defined cloudy days as those on which the C_{index} was below the 20th percentile across all stations (i.e., $C_{index} < 0.335$).

In January, when all days are considered irrespective of the day's C_{index} , the MBE was around 75 W m^{-2} but was around 100 W m^{-2} in April and July (Figs. 8a–d), which resulted in a positive MBE in air temperature (Figs. 8e–h) and 5-cm soil temperature (Figs. 8i–l). The HRRR captured the larger amplitude in the diurnal temperature variability on the subset of

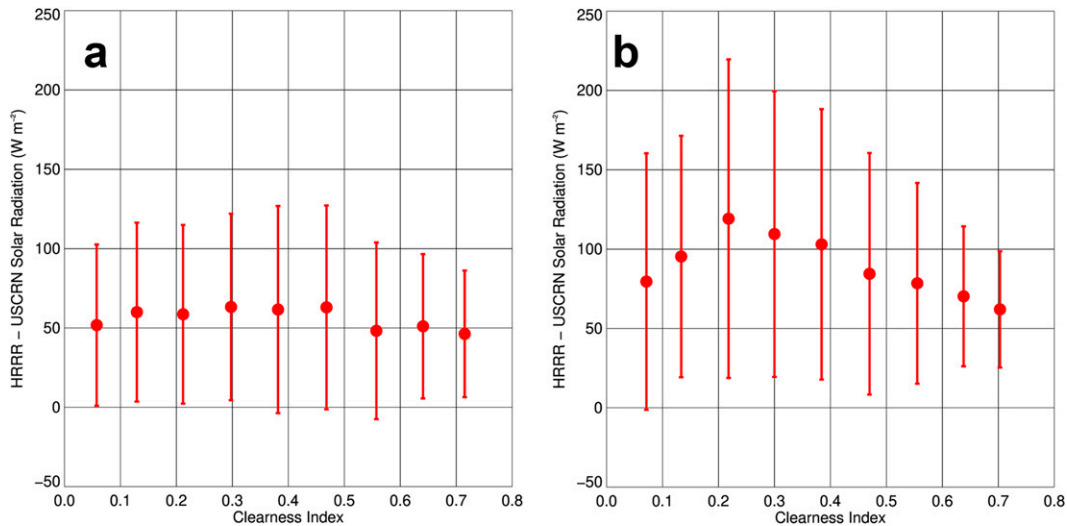


FIG. 9. Mean differences (filled red circles) ± 1 standard deviation (bars) between the HRRR 18-h forecast and USCRN SW_d as a function of the C_{index} in (a) January and (b) July. Note the differences in SW_d are only computed for hours in which the observed SW_d is positive.

clear days but overestimated minimum temperatures by as much as 3°C in the 18-h forecast which happened in October. Cloudy days had the largest overestimates in SW_d , with an MBE in the afternoon as large as 200 W m^{-2} in July in both the 1- and 18-h forecasts. Furthermore, the 18-h forecast more significantly overestimated SW_d during the early-midmorning on cloudy days as compared with the 1-h forecast.

As expected, the hourly averaged 5-cm soil moisture showed little diurnal variability in the observations and HRRR forecasts (Figs. 8m–p). When all days were considered, the HRRR underestimated soil moisture by on average 0.06 and $0.05\text{ m}^3\text{ m}^{-3}$ in January and April, respectively, whereas the offset between the model and observations was negligible in July and October. On the subset of clear days, the HRRR overestimated soil moisture by on average 0.07 and $0.03\text{ m}^3\text{ m}^{-3}$ in July and October, respectively. The MBE in the soil moisture was lower for cloudy days than for clear days. However, we note that, despite the relatively small offsets between the modeled and observed soil moisture, the standard deviations in the mean diurnal cycle of soil moisture were nontrivial and typically $\sim 0.12\text{ m}^3\text{ m}^{-3}$ in the USCRN observations, but $\sim 0.06\text{ m}^3\text{ m}^{-3}$ in the HRRR 1- and 18-h forecasts.

c. Potential sources of errors in the HRRR

1) ROLE OF ERRORS IN INCOMING SHORTWAVE RADIATION ON MODEL PERFORMANCE

As noted in section 4a, the HRRR generally predicts too much incoming shortwave radiation which generally correlates with a positive MBE in temperature. We investigated the errors in incoming shortwave radiation using the C_{index} described in section 3 and found an inverse relationship between the MBE in SW_d (taken only for when $SW_d > 0\text{ W m}^{-2}$) and the C_{index} . Over the entire 1-yr period, the MBE in SW_d was

around 40 W m^{-2} for days with a low C_{index} (i.e., cloudy days), but was around 15 W m^{-2} on days with a high C_{index} (i.e., clear days and days with optically thin clouds) and showed marked seasonal differences. In January, when the C_{index} was ~ 0.1 , the MBE for SW_d was $\sim 60\text{ W m}^{-2}$ and decreased only slightly as a function of C_{index} (Fig. 9a). In July, the MBE in SW_d ranged from 100 to 120 W m^{-2} on days with a low C_{index} to about 60 W m^{-2} on days with a high C_{index} (Fig. 9b). These results were consistent with those for other forecast hours (not shown), and the results agree with findings from Min et al. (2021), who similarly investigated the relationship between a clear-sky index and the difference between the observed and modeled incoming shortwave radiation. Min et al. (2021) noted that days with low C_{index} were associated with individual stratocumulus and cumulonimbus clouds which, due to their scale, were unable to be resolved by the HRRR.

To understand the radiation errors in more detail, we investigated the relationship between the HRRR's SW_d MBE and the within-hour variability in the observed SW_d to determine if the HRRR performed better when the observed SW_d variability was small (e.g., cases with stratus or stratocumulus clouds) compared with when the observed SW_d variability was large (e.g., cases with broken clouds or cumulus clouds). As the USCRN reports the 5-min values of SW_d , we used these values to compute the standard deviation in SW_d for each hour for each USCRN station. We found a statistically significant, positive relationship between the SW_d MBE and observed SW_d standard deviation that became weaker as a function of forecast hour in January (Fig. 10). In July, there was also a positive relationship between the SW_d MBE and the within-hour variability in the observed SW_d for all forecast hours. Similarly, the strength of this relationship was strongest for the 1-h forecast ($r = 0.29$, $p < 0.01$). The larger errors in SW_d for the cases with larger within-hour SW_d variability suggest the HRRR has difficulty resolving instances

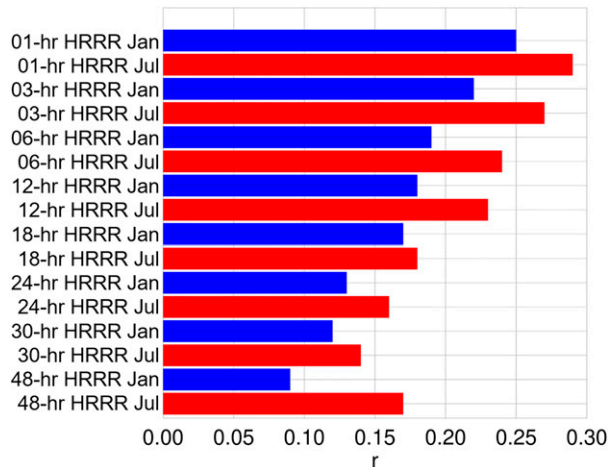


FIG. 10. Pearson correlation coefficient (r) for the relationship between the difference between the HRRR forecast and USCRN SW_d as a function of the observed SW_d standard deviation for the eight different forecast hours. All values are significant at the 0.01 confidence level (i.e., $p < 0.01$).

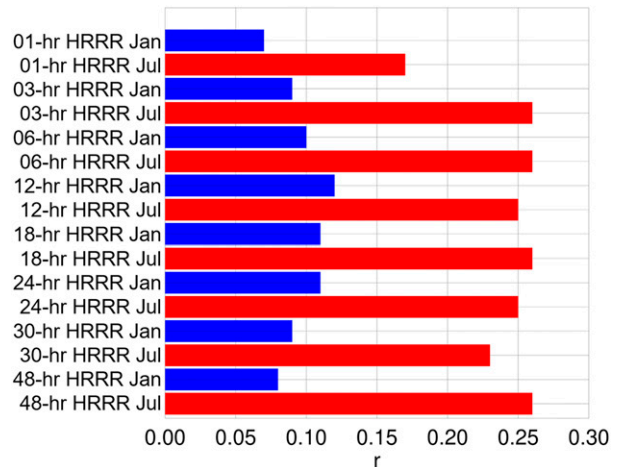


FIG. 11. Pearson correlation coefficient between the difference in modeled and observed SW_d vs the modeled and observed T_a for the eight forecast periods investigated in the present study and only for cases when the observed SW_d is positive. All values are significant at the 0.01 confidence level (i.e., $p < 0.01$).

with broken or scattered clouds which occur on spatial and temporal scales that are too small to be resolved by the model.

We also found a statistically significant positive relationship between the SW_d MBE and the T_a MBE that was stronger in the summer than during the winter (Fig. 11). In July the relationship between these variables for the 18-h HRRR forecast was 0.26 ($p < 0.01$), whereas in January it was 0.11 ($p < 0.01$). These results are consistent with Min et al. (2021) who similarly observed stronger correlations between SW_d and T_a during the warm season than during the cold season. The smaller correlations during the cold season are likely related to other factors (e.g., snow cover increasing the albedo, frozen soils, etc.) that reduce the importance of incoming solar radiation on modeled temperature.

2) ROLE OF DIFFERENT SOIL MOISTURE AMOUNTS ON MODEL PERFORMANCE

As noted in section 4a, the HRRR generally overestimated soil moisture when the observed soil moisture was $< 0.2 \text{ m}^3 \text{ m}^{-3}$ but underestimated the soil moisture when the observed soil moisture exceeded this value. Consequently, there was a strong inverse relationship between the observed soil moisture and the MBE in the 18-h HRRR forecast that is present both in January (Fig. 12a) and in July (Fig. 12b), where $r = -0.839$ ($p < 0.001$) and $r = -0.854$ ($p < 0.001$), respectively, as well as for the other forecast hours (not shown). Additionally, the slope of the relationship between the soil moisture MBE in the 18-h forecast and the observed soil moisture was -0.70 and -0.64 in January and July, respectively, with similar results for the other forecast hours. The strong inverse linear relationship implies a possible issue with the representation of soil moisture conductivity in the HRRR.

We also found a relationship between the temperature MBE and the observed soil moisture when all times of day were considered. In July, there was a positive relationship between HRRR’s air temperature MBE and the observed 5-cm soil moisture at the USCRN stations. The relationship was stronger for the earlier forecast hours than for the later forecast hours. For example, in the 1-h forecast, r was 0.19 ($p < 0.01$), but in the 48-h forecast, r was 0.07 ($p < 0.01$) (Fig. 13). In January, we found no relationship between the MBE in the 1-h HRRR’s air temperature and observed soil moisture ($r = 0.04, p = 0.13$), whereas the later forecast periods had a statistically significant inverse relationship, as $r = -0.22$ ($p < 0.01$) in the 48-h forecast.

Because the daytime 2-m observations of temperature and dewpoint temperature are used to make adjustments to soil temperature and moisture in the HRRR, we also evaluated the relationship between the HRRR’s temperature MBE and observed soil moisture but only for the daytime, i.e., when $SW_d > 0$ (Fig. 14). We found similar results to the results reported above, the exception was the relationship in the 48-h forecast in July when the relationship between these variables shifted from a weak positive relationship (cf. Fig. 13f; $r = 0.07, p < 0.01$) to a stronger negative relationship (Fig. 14f; $r = -0.38, p < 0.01$). Despite there generally being statistically significant relationships here, the lack of a strong correlation between the temperature errors and soil moisture suggests other factors besides these contribute to the HRRR’s temperature biases, e.g., errors in the HRRR’s heat fluxes which have been reported as a source of bias in previous work (e.g., Lee et al. 2019).

d. HRRR performance by NCEI region

When averaged across the year, nearly all USCRN stations had a positive MBE in SW_d (Fig. 15a). Stations in the Southeast had the largest RMSE (36 W m^{-2}) and the lowest R^2 and

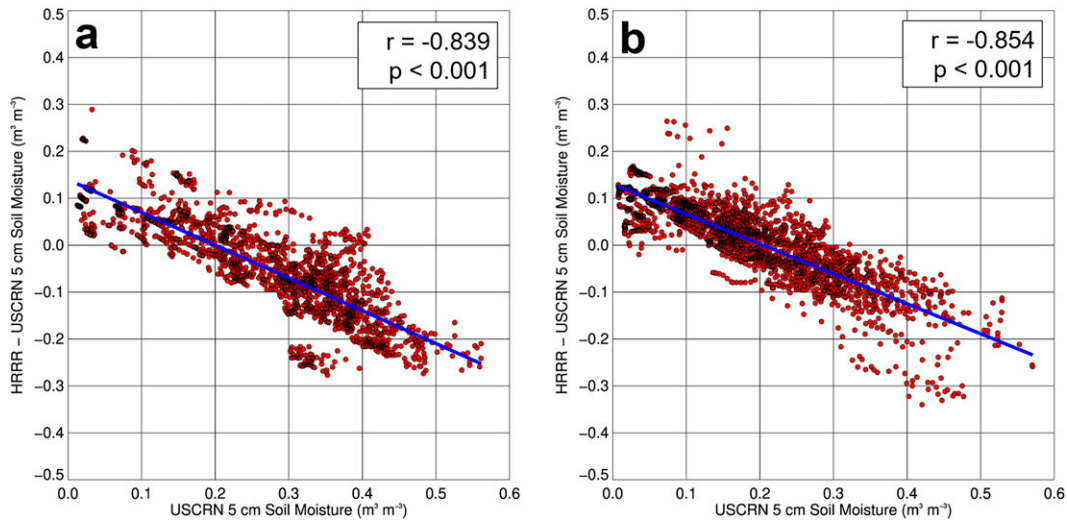


FIG. 12. Modeled minus observed 5-cm soil moisture as a function of 5-cm soil moisture observed at all USCRN stations for the HRRR 18-h forecast in (a) January and (b) July. The blue line shows the line of best fit, and the values for r and p are shown in the top right of each panel.

m_b of all regions in the 18-h HRRR forecast (Fig. 16). This was also the case in the other forecast hours (not shown). Air temperature correspondingly showed a positive MBE across the United States, although a few USCRN stations in the Southwest, Northern Rockies and Plains and along the East Coast had a slightly negative MBE (Fig. 15b). For the 18-h forecast, the MBE was $<1^\circ\text{C}$ across all USCRN stations, and the RMSE ranged from 1.19°C for the Northeast to 1.66°C for stations in the Northern Rockies and Plains. Surface temperature generally followed this pattern, although T_s had slightly lower R^2 and m_b and a higher RMSE than T_a . The exception was the West where T_s had a more pronounced negative MBE than T_a (Fig. 15c), resulting in an MBE of -1.0°C in the 18-h forecast for stations in the West.

When averaged across the entire year, 5- and 10-cm soil temperatures generally showed a negative MBE across all regions (Figs. 15d,e) but there was significant variability in these values as indicated by the large standard deviations at the 5- and 10-cm depths and RMSEs that were about 2–3 times larger than the RMSEs for T_a . The larger RMSEs for soil temperature as compared with air temperature were apparent not only in the 18-h HRRR forecast (Fig. 16) but for other forecast periods as well (not shown).

The model overestimated precipitation (Fig. 15f), with the largest overestimates occurring over the Southeast. The overestimates in this particular region of the United States may be due to e.g., the highly spatial nature of warm-season convection and errors associated with subgrid-scale precipitation gradients compared to gauge measurements. Model performance was generally best over the western United States; R^2 between the observations and 18-h forecast was 0.66 and 0.53 over the Northwest and West, respectively, but was 0.05 for the Southeast and South (Fig. 16). It is important to note that our analyses did not show any consistent evidence of undercatch at the USCRN stations that more routinely experience frozen

precipitation. However, we found smaller biases in precipitation at stations in the Northern Rockies and Plains where frozen precipitation is more common than in other regions. The USCRN goes to great lengths to ensure that snow is accurately measured as water equivalent of snow using a small double-fence intercomparison reference coupled with a heated weighing gauge. The wind fencing around the precipitation gauge, including a single-alter shield, ensures the most accurate measurement of precipitation (e.g., Kochendorfer et al. 2022).

As we found for precipitation, we found that the HRRR had strong regional differences in soil moisture, showing a positive MBE in the western United States and negative MBE in the eastern United States (Figs. 15g,h and 16). In the Southwest, Northwest, and West (where soil moisture conditions are generally well below $0.2\text{ m}^3\text{ m}^{-3}$), the HRRR overestimated 5-cm soil moisture in the 18-h forecast by 0.07 , 0.06 , and $0.07\text{ m}^3\text{ m}^{-3}$, respectively, whereas the HRRR underestimated 5-cm soil moisture by $0.09\text{ m}^3\text{ m}^{-3}$ over the Ohio Valley. We also found significant variability in the model performance across the different regions. The model performance was best in the Upper Midwest and across the Northern Rockies and Plains, where m_b was 0.40 and 0.63, respectively, and R^2 was 0.75 and 0.74, respectively. In contrast, the Southeast had the poorest agreement between the modeled and observed 5-cm soil moisture, as m_b and R^2 were 0.09 and 0.02, respectively. These results were consistent for the 10-cm soil moisture values, as well as for the other forecast hours (not shown). We suspect the poor agreement in soil moisture occurs due to the aforementioned spatial nature of warm-season convection over this region which results in fine-scale precipitation gradients.

e. HRRR performance by land cover type

We found a positive MBE in air temperature across all land cover types. The positive MBE coincided with a positive SW_d MBE (Fig. 17). The MBE in SW_d ranged from 20 W m^{-2} over

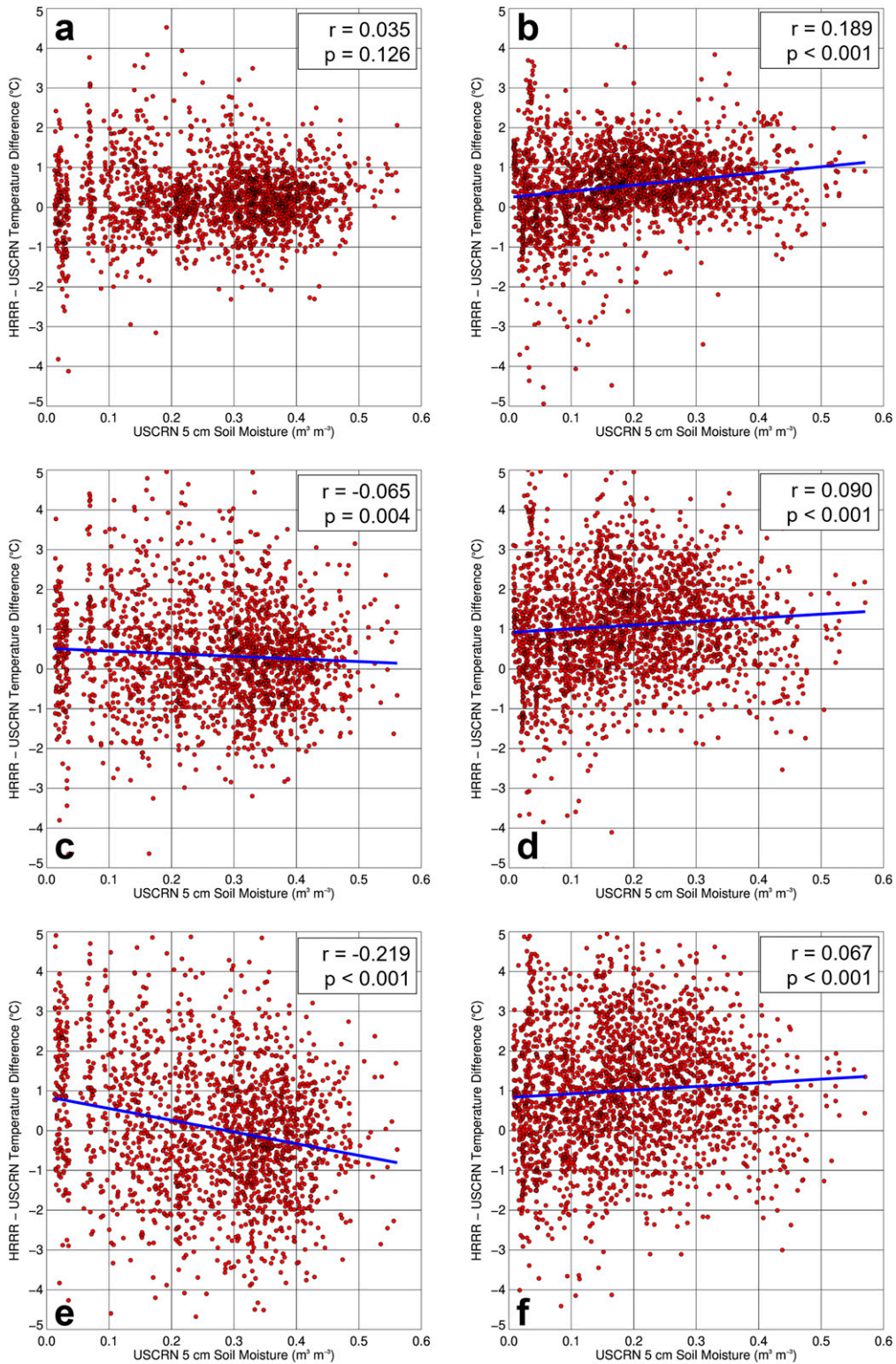


FIG. 13. Modeled minus observed air temperature as a function of 5-cm soil moisture observed at all USCRN stations for the HRRR 1-h forecast in (a) January and (b) July. (c),(d) As in (a) and (b), but for the HRRR 18-h forecast. (e),(f) As in (a) and (b), but for the HRRR 48-h forecast. The blue line shows the line of best fit for cases in which there is a statistically significant relationship. The values for r and p are shown in the top right of each panel.

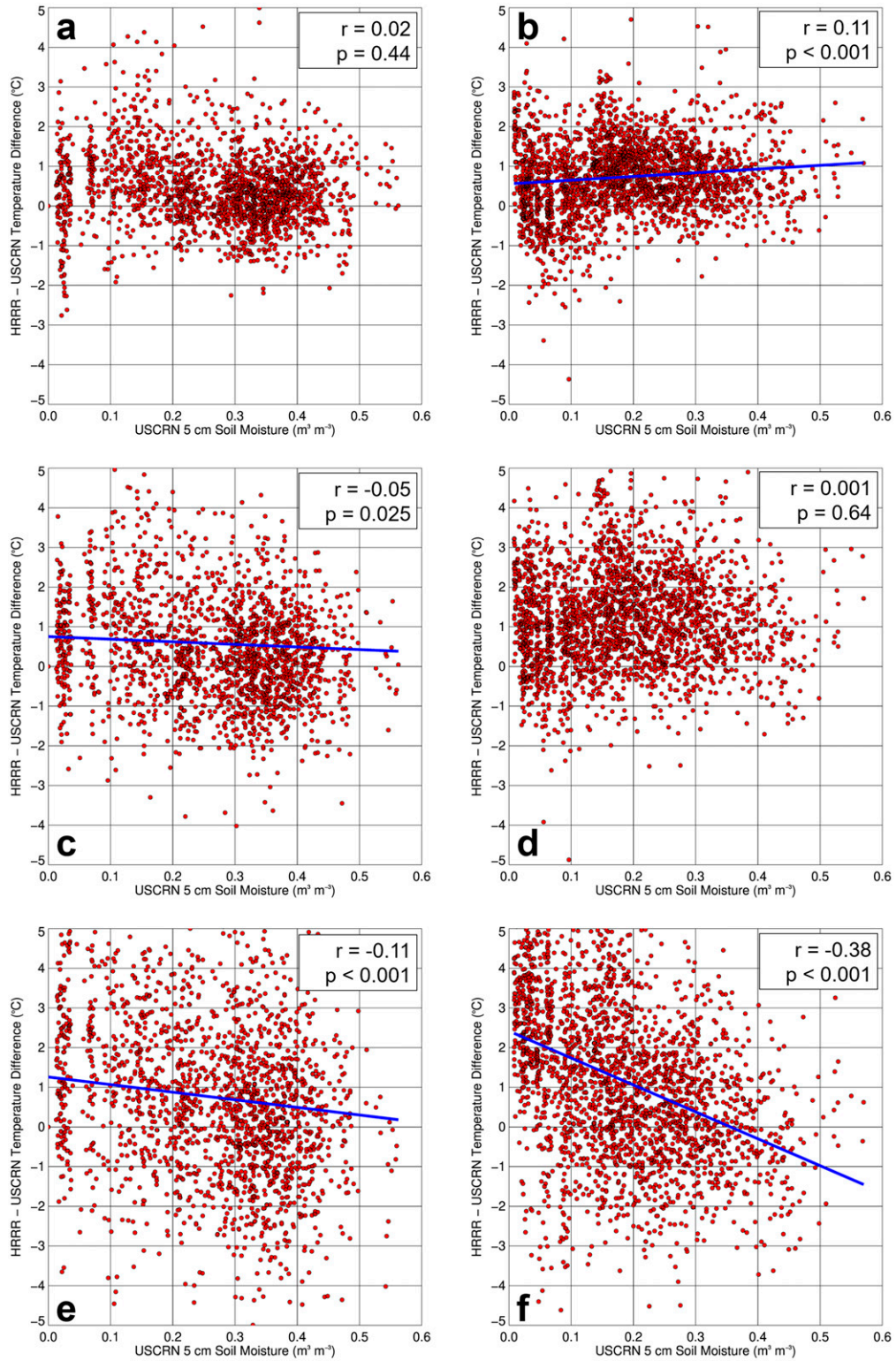


FIG. 14. As in Fig. 13, but only for cases when $\text{SW}_d > 0$.

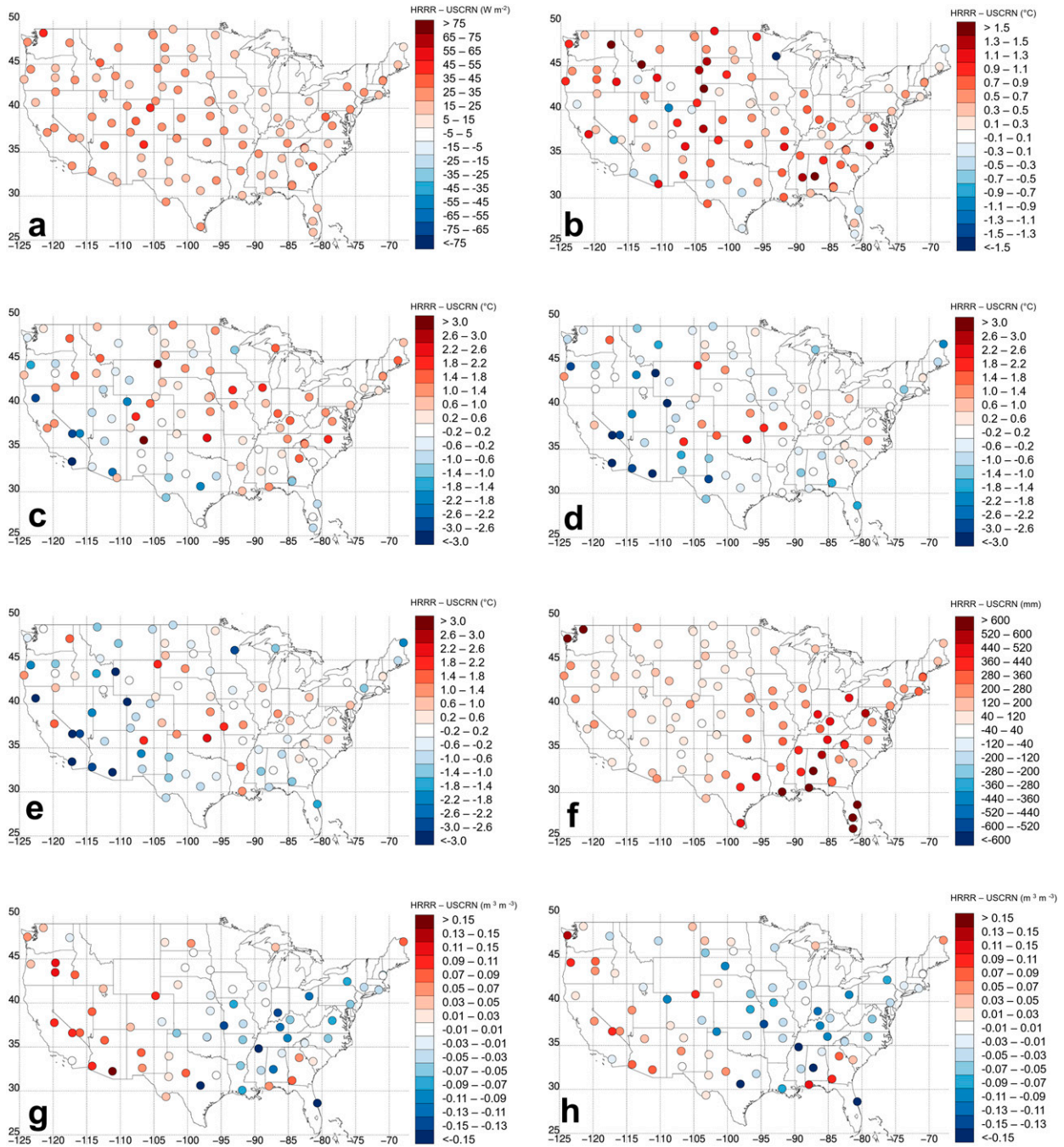


FIG. 15. Spatial variability in the difference between the 18-h HRRR forecast and the USCRN observations over the contiguous United States for (a) SW_d , (b) T_a , (c) T_s , (d) ST_{05} , (e) ST_{10} , (f) P , (g) SM_{05} , and (h) SM_{10} . The values represent the mean daily differences, except for P , in which the daily totals for the entire year are used both for the USCRN observations and HRRR model output prior to computing the differences.

the savanna to $28 W m^{-2}$ over the grassland land cover type. All land cover classifications had an m_b and R^2 of ~ 1 and ~ 0.9 , respectively, for the relationships between the modeled and observed SW_d .

Surface temperatures had a mean positive MBE for forest, savanna, grassland, and cropland, but a mean negative MBE

for shrubland. As was the case when classifying the stations by region, the standard deviations were about twice as large for surface temperature as they were for air temperature. Also, all land cover types had fairly similar m_b , R^2 , and RMSE values. Soil temperatures had a negative MBE across the different land cover types, but there was overall good

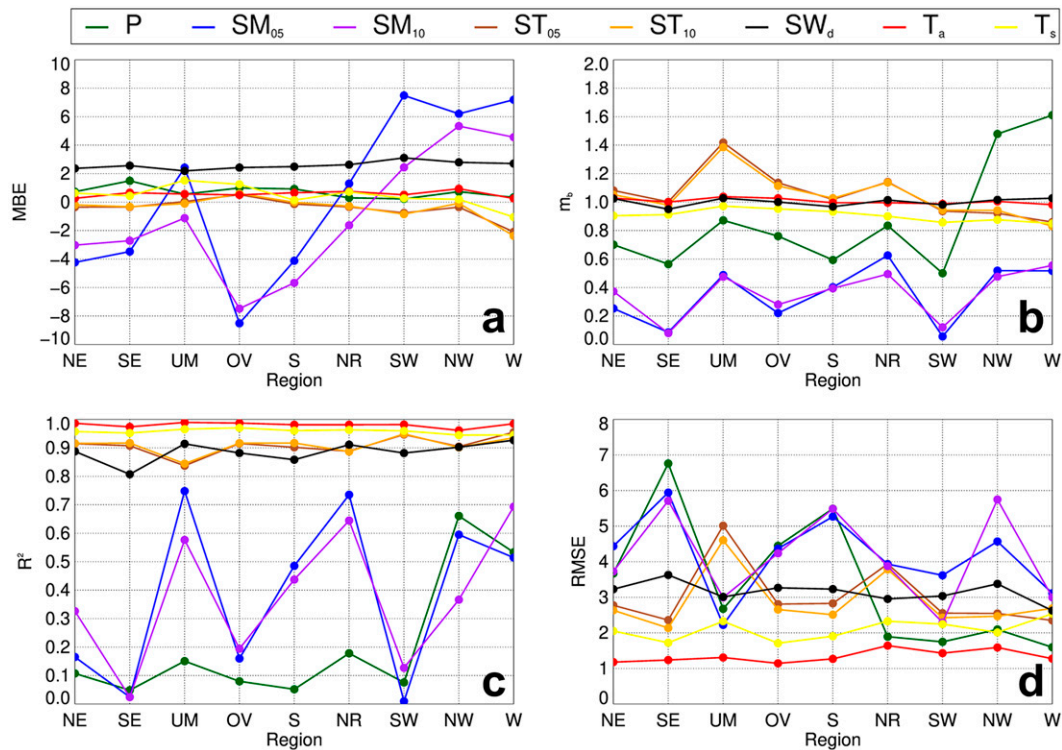


FIG. 16. (a) MBE, (b) m_b , (c) R^2 , and (d) RMSE for P (green line), SM_{05} (blue line), SM_{10} (purple line), ST_{05} (brown line), ST_{10} (orange line), SW_d (black line), T_a (red line), and T_s (yellow line) for the nine NCEI regions identified in Fig. 1a: Northeast (NE), Southeast (SE), Upper Midwest (UM), Ohio Valley (OV), South (S), Northern Rockies and Plains (NR), Southwest (SW), Northwest (NW), and West (W). To show all lines on the same graphs, the MBE and RMSE for SM_{05} and SM_{10} have been multiplied by 100, and the MBE and RMSE for SW_d has been multiplied by 0.1. The units of MBE and RMSE for precipitation, soil moisture, temperature, and radiation are mm, $m^3 m^{-3}$, $^{\circ}C$, and $W m^{-2}$, respectively.

agreement between the model and observations as evident by $R^2 > 0.9$. RMSE was largest for the cropland land cover type for all depths.

The HRRR's largest overestimates in precipitation were for the cropland land cover type, and the HRRR best predicted precipitation over the shrubland land cover type. The mean soil moisture MBE values at 5 cm were largest over the shrubland. The agreement between the observations and HRRR was best over the savanna where m_b and R^2 were 0.40 and 0.68 at the 5-cm level. The results were generally consistent at 10 cm, although we note that m_b was slightly larger over the shrubland (0.41) than it was for the savanna (0.38).

5. Summary and conclusions

We presented the first known study evaluating the performance of the HRRR's subsurface moisture and temperature fields over the contiguous United States, using observations obtained from the USCRN between 1 January and 31 December 2021. Over this entire period, the HRRR best simulated near-surface air temperature with RMSEs ranging from $1^{\circ}C$ at the 1-h forecast to around $2^{\circ}C$ in the 48-h forecast. RMSEs were generally larger for surface and subsurface temperatures than

they were for air temperature. Importantly, soil temperatures had larger values of RMSE when the observed soil temperatures were below freezing.

We next investigated possible error sources in the HRRR by studying the role of HRRR's performance under different radiative conditions and different soil moisture regimes. We found that the downwelling shortwave radiation errors were generally largest under cloudy skies (i.e., low C_{index}). The radiation errors were larger for cases with broken clouds than for cases with stratiform clouds, and there was a positive relationship between the MBE in incoming shortwave radiation and air temperature that was stronger in the summer than in the winter. We also found that the HRRR overestimated soil moisture for dry soils (i.e., $<0.2 m^3 m^{-3}$) but generally underestimated soil moisture when the soil moisture exceeded $0.2 m^3 m^{-3}$. Temperature errors as a function of soil moisture varied by forecast hour and season. There was a positive relationship between the MBE in the HRRR forecast of air temperature and observed soil moisture in July, which may be related to a larger amount of sensible heating in the model caused by the drier biases for moist soils. Soil moisture may also contribute to some of the warm season biases when land cover processes (e.g., evaporation, surface roughness, etc.) and associated atmospheric landcover interactions are most prominent.

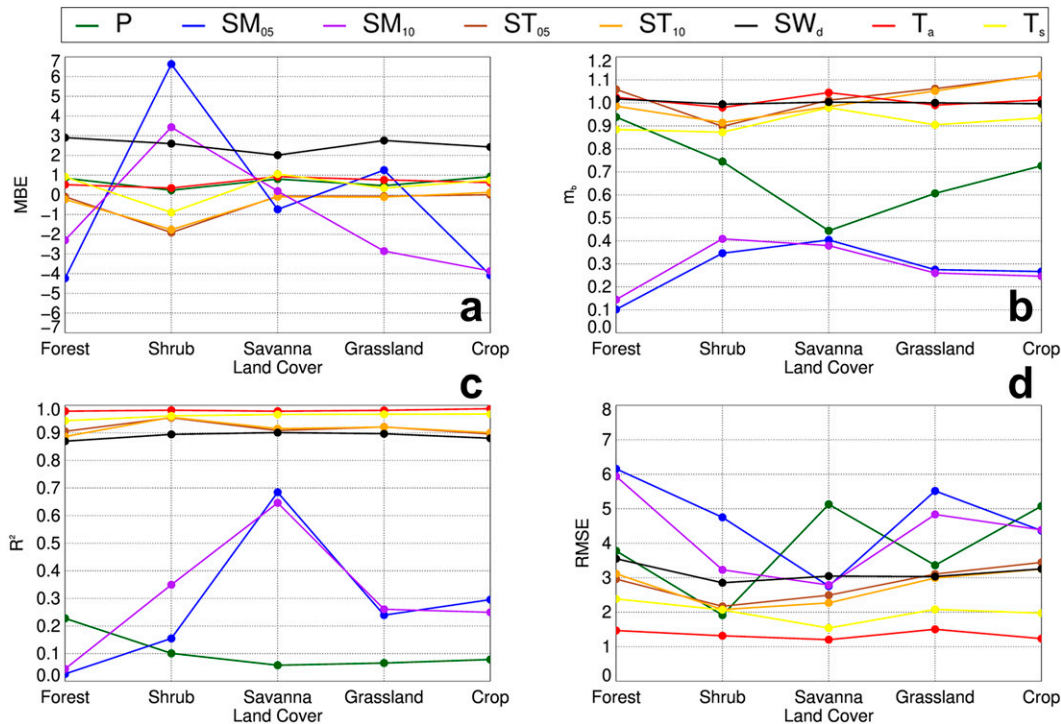


FIG. 17. (a) MBE, (b) m_b , (c) R^2 , and (d) RMSE for P (green line), SM_{05} (blue line), SM_{10} (purple line), ST_{05} (brown line), ST_{10} (orange line), SW_d (black line), T_a (red line), and T_s (yellow line) for the different land cover types. To show all lines on the same graphs, the MBE and RMSE for SM_{05} and SM_{10} have been multiplied by 100, and the MBE and RMSE for SW_d have been multiplied by 0.1. The units of MBE and RMSE for precipitation, soil moisture, temperature, and radiation are mm, $m^3 m^{-3}$, $^{\circ}C$, and $W m^{-2}$, respectively.

We further evaluated the HRRR’s performance for different geographic regions of the United States and land cover types. The most notable regional MBEs across were present in soil moisture, where the 5- and 10-cm depths exhibited a positive MBE in the drier western United States and a negative MBE in the wetter eastern United States. Therefore, there existed a relatively tight linear relationship between the observed soil moisture and the soil moisture MBE which implies a possible issue with the soil moisture conductivity in the HRRR. Overall, the HRRR best simulated soil moisture over the Upper Midwest as well as the Northern Rockies and Plains but did not simulate soil moisture well over the Southeast United States. When classifying the USCRN stations by land cover type, we found that the HRRR overestimated incoming shortwave radiation and air temperature, and we found that the USCRN stations with the shrubland land cover classification had the largest errors in 5-cm soil moisture. However, it is important to reiterate caution when interpreting conclusions regarding the HRRR’s performance under certain conditions, e.g., 1) stations where the HRRR land cover type is not entirely representative of the land cover type at the USCRN station, e.g., USCRN stations at which the RUC LSM used in the HRRR assigns the forest land cover type; 2) USCRN stations located in mountainous areas, e.g., over the western United States, and/or stations with subgrid-scale variations in land cover type; and 3) instances in which the HRRR has partial snow cover.

Despite the aforementioned caveats, the results from this work build upon previous research that has identified ways to improve upon recent operational versions of the HRRR that have noted e.g., biases in surface fluxes (Lee et al. 2019), biases in PBL observations of temperature and wind (e.g., Fovell and Gallagher 2020), biases within mesonets (Min et al. 2021), etc. Careful consideration of these biases, coupled with the development of and modifications to new surface-layer parameterization schemes (e.g., Sorbjan 2017; Lee and Buban 2020; Lee et al. 2021; Greene et al. 2022; Wulfmeyer et al. 2023; Lee and Meyers 2023), will lead to improvements in the LSMs used in high-resolution weather forecasting models.

Acknowledgments. We gratefully acknowledge the engineers and technicians from the NOAA Air Resources Laboratory Atmospheric Turbulence and Diffusion Division for their efforts installing and maintaining the USCRN to ensure a dataset that is robust and of high quality, as well as the scientists and data specialists from NOAA’s National Centers for Environmental Information for the data ingest, quality control, access, and archive work that allows for the USCRN data to be easily available to users in a timely manner. We also thank Dr. Stan Benjamin and Dr. Joe Olson of the NOAA Global Systems Laboratory for helpful insights on the HRRR’s data assimilation system, and we gratefully acknowledge the two anonymous reviewers whose feedback led to an improved manuscript. This work was supported, in part, by

the NOAA Atmospheric Science for Renewable Energy (ASRE) program and by NOAA's Cooperative Institute for Satellite Earth System Studies under Cooperative Agreement NA19NES4320002. Finally, we note that the results and conclusions of this study, as well as any views expressed herein, are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

Data availability statement. The USCRN datasets used in this study are available from <https://www.ncei.noaa.gov/access/crn/qcdatasets.html>, and the HRRR forecasts can be accessed through the Amazon Web Services HRRR archive at <https://registry.opendata.aws/noaa-hrrr-pds/>.

REFERENCES

- Alexander, G. A., H. A. Holmes, X. Sun, D. Caputi, I. C. Faloona, and H. J. Oldroyd, 2022: Simulating land-atmosphere coupling in the Central Valley, California: Investigating soil moisture impacts on boundary layer properties. *Agric. For. Meteorol.*, **317**, 108898, <https://doi.org/10.1016/j.agrformet.2022.108898>.
- Augustine, J. A., J. J. DeLuisi, and C. N. Long, 2000: SURFRAD—A national surface radiation budget network for atmospheric research. *Bull. Amer. Meteor. Soc.*, **81**, 2341–2358, [https://doi.org/10.1175/1520-0477\(2000\)081<2341:SANSRB>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2).
- Bell, J. E., and Coauthors, 2013: U.S. Climate Reference Network soil moisture and temperature observations. *J. Hydrometeorol.*, **14**, 977–988, <https://doi.org/10.1175/JHM-D-12-0146.1>.
- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- , T. G. Smirnova, E. P. James, L.-F. Lin, M. Hu, D. D. Turner, and S. He, 2022: Land-snow data assimilation including a moderately coupled initialization method applied to NWP. *J. Hydrometeorol.*, **23**, 825–845, <https://doi.org/10.1175/JHM-D-21-0198.1>.
- Blaylock, B. K., J. D. Horel, and S. T. Liston, 2017: Cloud archiving and data mining of High-Resolution Rapid Refresh forecast model output. *Comput. Geosci.*, **109**, 43–50, <https://doi.org/10.1016/j.cageo.2017.08.005>.
- Buban, M. S., T. R. Lee, and C. B. Baker, 2020: A comparison of the U.S. climate reference network precipitation data to the Parameter-Elevation Regressions on Independent Slopes Model (PRISM). *J. Hydrometeorol.*, **21**, 2391–2400, <https://doi.org/10.1175/JHM-D-19-0232.1>.
- Diamond, H. J., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. *Bull. Amer. Meteor. Soc.*, **94**, 485–498, <https://doi.org/10.1175/BAMS-D-12-00170.1>.
- Dirmeyer, P. A., and Coauthors, 2012: Evidence for enhanced land-atmosphere feedback in a warming climate. *J. Hydrometeorol.*, **13**, 981–995, <https://doi.org/10.1175/JHM-D-11-0104.1>.
- Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, <https://doi.org/10.1175/WAF-D-21-0151.1>.
- Ek, M. B., and A. A. M. Holtslag, 2004: Influence of soil moisture on boundary layer cloud development. *J. Hydrometeorol.*, **5**, 86–99, [https://doi.org/10.1175/1525-7541\(2004\)005<0086:IOSMOB>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0086:IOSMOB>2.0.CO;2).
- , K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, <https://doi.org/10.1029/2002JD003296>.
- Fan, Y., H. M. van den Dool, and W. Wu, 2011: Verification and intercomparison of multimodel simulated land surface hydrological datasets over the United States. *J. Hydrometeorol.*, **12**, 531–555, <https://doi.org/10.1175/2011JHM1317.1>.
- Fovell, R. G., and A. Gallagher, 2020: Boundary layer and surface verification of the High-Resolution Rapid Refresh, version 3. *Wea. Forecasting*, **35**, 2255–2278, <https://doi.org/10.1175/WAF-D-20-0101.1>.
- Gallo, K., and P. Krishnan, 2022: Evaluation of the bias in the use of clear-sky compared with all-sky observations of monthly and annual daytime land surface temperature. *J. Appl. Meteor. Climatol.*, **61**, 1485–1495, <https://doi.org/10.1175/JAMC-D-21-0240.1>.
- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, <https://doi.org/10.1175/2009WAF2222269.1>.
- Greene, B. R., S. T. Kral, P. B. Chilson, and J. Reuder, 2022: Gradient-based turbulence estimates from multicopter profiles in the Arctic stable boundary layer. *Bound.-Layer Meteorol.*, **183**, 321–353, <https://doi.org/10.1007/s10546-022-00693-x>.
- He, S., T. G. Smirnova, and S. G. Benjamin, 2021: Single-column validation of a snow subgrid parameterization in the rapid update cycle land-surface model (RUC LSM). *Water Resour. Res.*, **57**, e2021WR029955, <https://doi.org/10.1029/2021WR029955>.
- Hubbard, K. G., X. Lin, C. B. Baker, and B. Sun, 2004: Air temperature comparison between MMTS and the USCRN temperature systems. *J. Atmos. Oceanic Technol.*, **21**, 1590–1597, [https://doi.org/10.1175/1520-0426\(2004\)021<1590:ATCBTM>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<1590:ATCBTM>2.0.CO;2).
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating High-Resolution Rapid Refresh model. *Wea. Forecasting*, **28**, 921–939, <https://doi.org/10.1175/WAF-D-12-00085.1>.
- James, E. P., and Coauthors 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, <https://doi.org/10.1175/WAF-D-21-0130.1>.
- Kantha Rao, B., and V. Rakesh, 2019: Evaluation of WRF-simulated multilevel soil moisture, 2-m air temperature, and 2-m relative humidity against in situ observations in India. *Pure Appl. Geophys.*, **176**, 1807–1826, <https://doi.org/10.1007/s00024-018-2022-7>.
- Kochendorfer, J., and Coauthors, 2022: How well are we measuring snow post-SPICE? *Bull. Amer. Meteor. Soc.*, **103**, E370–E388, <https://doi.org/10.1175/BAMS-D-20-0228.1>.
- Krishnan, P., J. Kochendorfer, E. J. Dumas, P. C. Guillevic, C. Bruce Baker, T. P. Meyers, and B. Martos, 2015: Comparison of in-situ, aircraft, and satellite land surface temperature measurements over a NOAA climate reference network site.

- Remote Sens. Environ.*, **165**, 249–264, <https://doi.org/10.1016/j.rse.2015.05.011>.
- Lee, C. B., J.-C. Kim, M. Belorid, and P. Zhao, 2016: Performance evaluation of four different land surface models in WRF. *Asian J. Atmos. Environ.*, **10**, 42–50, <https://doi.org/10.5572/ajae.2016.10.1.042>.
- Lee, T. R., and M. Buban, 2020: Evaluation of Monin–Obukhov and bulk Richardson parameterizations for surface–atmosphere exchange. *J. Appl. Meteor. Climatol.*, **59**, 1091–1107, <https://doi.org/10.1175/JAMC-D-19-0057.1>.
- , and T. P. Meyers, 2023: New parameterizations of turbulence statistics for the atmospheric surface layer. *Mon. Wea. Rev.*, **151**, 85–103, <https://doi.org/10.1175/MWR-D-22-0071.1>.
- , S. F. J. De Wekker, S. Pal, A. E. Andrews, and J. Kofler, 2015: Meteorological controls on the diurnal variability of carbon monoxide mixing ratio at a mountaintop monitoring site in the Appalachian Mountains. *Tellus*, **67B**, 25659, <https://doi.org/10.3402/tellusb.v67.25659>.
- , M. Buban, M. A. Palecki, R. D. Leeper, H. J. Diamond, E. Dumas, T. P. Meyers, and C. B. Baker, 2018: Great American eclipse data may fine-tune weather forecasts. *Eos*, **99**, <https://doi.org/10.1029/2018EO103931>.
- , —, D. D. Turner, T. P. Meyers, and C. B. Baker, 2019: Evaluation of the High-Resolution Rapid Refresh (HRRR) Model using near-surface meteorological and flux observations from northern Alabama. *Wea. Forecasting*, **34**, 635–663, <https://doi.org/10.1175/WAF-D-18-0184.1>.
- , —, and T. P. Meyers, 2021: Application of bulk Richardson parameterizations of surface fluxes to heterogeneous land surfaces. *Mon. Wea. Rev.*, **149**, 3243–3264, <https://doi.org/10.1175/MWR-D-21-0047.1>.
- Leeper, R. D., R. Mahmood, and A. I. Quintanar, 2009: Near surface atmospheric response to simulated changes in land-cover, vegetation fraction, and soil moisture over Western Kentucky. *Publ. Climatol.*, **62**, 1–41.
- , J. Rennie, and M. A. Palecki, 2015: Observational perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) network: Temperature and precipitation comparison. *J. Atmos. Oceanic Technol.*, **32**, 703–721, <https://doi.org/10.1175/JTECH-D-14-00172.1>.
- , J. E. Bell, C. Vines, and M. Palecki, 2017: An evaluation of the North American Regional Reanalysis simulated soil moisture conditions during the 2011–13 drought period. *J. Hydrometeorol.*, **18**, 515–527, <https://doi.org/10.1175/JHM-D-16-0132.1>.
- , B. Petersen, M. A. Palecki, and H. Diamond, 2021: Exploring the use of standardized soil moisture as a drought indicator. *J. Appl. Meteor. Climatol.*, **60**, 1021–1033, <https://doi.org/10.1175/JAMC-D-20-0275.1>.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Min, L., D. R. Fitzjarrald, Y. Du, B. E. J. Rose, J. Hong, and Q. Min, 2021: Exploring sources of surface bias in HRRR using New York state mesonet. *J. Geophys. Res. Atmos.*, **126**, e2021JD034989, <https://doi.org/10.1029/2021JD034989>.
- Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteorol.*, **112**, 1–31, <https://doi.org/10.1023/B:BOUND.0000020164.04146.98>.
- , and —, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, <https://doi.org/10.2151/jmsj.87.895>.
- Oke, T. R., 1987: *Boundary Layer Climates*. 2nd ed. Routledge Press, 464 pp.
- Olson, J. B., J. S. Kenyon, W. A. Angevine, J. M. Brown, M. Pagowski, and K. Suselj, 2019: A description of the MYNN-EDMF scheme and the coupling to other components in WRF–ARW. NOAA Tech. Memo. OAR GSD-61, 37 pp., <https://doi.org/10.25923/n9wm-be49>.
- , T. Smirnova, J. S. Kenyon, D. D. Turner, J. M. Brown, W. Zheng, and B. W. Green, 2021: A description of the MYNN surface-layer scheme. NOAA Tech. Memo. OAR GSL-67, 26 pp., <https://doi.org/10.25923/f6a8-bc75>.
- Otkin, J. A., M. C. Anderson, J. R. Mecikalski, and G. R. Diak, 2005: Validation of GOES-based insolation estimates using data from the U.S. climate reference network. *J. Hydrometeorol.*, **6**, 460–475, <https://doi.org/10.1175/JHM440.1>.
- Pielke, R. A., Sr., 2001: Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Rev. Geophys.*, **39**, 151–177, <https://doi.org/10.1029/1999RG000072>.
- Pleim, J. E., and A. Xiu, 2003: Development of a land surface model. Part II: Data assimilation. *J. Appl. Meteor.*, **42**, 1811–1822, [https://doi.org/10.1175/1520-0450\(2003\)042<1811:DOALSM>2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042<1811:DOALSM>2.0.CO;2).
- Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation testbed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, <https://doi.org/10.1175/BAMS-D-11-00052.1>.
- Rennie, J. J., M. A. Palecki, S. P. Heuser, and H. J. Diamond, 2021: Developing and validating heat exposure products using the U.S. climate reference network. *J. Appl. Meteor. Climatol.*, **60**, 543–558, <https://doi.org/10.1175/JAMC-D-20-0282.1>.
- Richards, L. A., 1931: Capillary conduction of liquids through porous mediums. *Physics*, **1**, 318–333, <https://doi.org/10.1063/1.1745010>.
- Santanello, J. A., Jr., P. Lawston, S. Kumar, and E. Dennis, 2019: Understanding the impacts of soil moisture initial conditions on NWP in the context of land–atmosphere coupling. *J. Hydrometeorol.*, **20**, 793–819, <https://doi.org/10.1175/JHM-D-18-0186.1>.
- Segal, M., and R. W. Arritt, 1992: Nonclassical mesoscale circulations caused by surface sensible heat-flux gradients. *Bull. Amer. Meteor. Soc.*, **73**, 1593–1604, [https://doi.org/10.1175/1520-0477\(1992\)073<1593:NMCCBS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<1593:NMCCBS>2.0.CO;2).
- Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Mon. Wea. Rev.*, **125**, 1870–1884, [https://doi.org/10.1175/1520-0493\(1997\)125<1870:PODSMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1870:PODSMC>2.0.CO;2).
- , —, —, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, <https://doi.org/10.1175/MWR-D-15-0198.1>.
- Sorbjan, Z., 2017: Assessment of gradient-based similarity functions in the stable boundary layer derived from a large-eddy simulation. *Bound.-Layer Meteorol.*, **163**, 375–392, <https://doi.org/10.1007/s10546-017-0234-5>.
- Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic, 666 pp.
- Sun, B., C. Bruce Baker, T. R. Karl, and M. D. Gifford, 2005: A comparative study of ASOS and USCRN temperature measurements. *J. Atmos. Oceanic Technol.*, **22**, 679–686, <https://doi.org/10.1175/JTECH1752.1>.

- Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, <https://doi.org/10.1175/JAS-D-13-0305.1>.
- Tian, J., Y. Zhang, S. A. Klein, R. Öktem, and L. Wang, 2022: How does land cover and its heterogeneity length scales affect the formation of summertime shallow cumulus clouds in observations from the U.S. Southern Great Plains. *Geophys. Res. Lett.*, **49**, e2021GL097070, <https://doi.org/10.1029/2021GL097070>.
- Viterbo, P., A. Beljaars, J.-F. Mahfouf, and J. Teixeira, 1999: The representation of soil moisture freezing and its impact on the stable boundary layer. *Quart. J. Roy. Meteor. Soc.*, **125**, 2401–2426, <https://doi.org/10.1002/qj.49712555904>.
- Wagner, T. J., P. M. Klein, and D. D. Turner, 2019: A new generation of ground-based mobile platforms for active and passive profiling of the boundary layer. *Bull. Amer. Meteor. Soc.*, **100**, 137–153, <https://doi.org/10.1175/BAMS-D-17-0165.1>.
- Whiteman, C. D., and K. J. Allwine, 1986: Extraterrestrial solar radiation on inclined surfaces. *Environ. Software*, **1**, 164–169, [https://doi.org/10.1016/0266-9838\(86\)90020-1](https://doi.org/10.1016/0266-9838(86)90020-1).
- , X. Bian, and S. Zhong, 1999: Wintertime evolution of the temperature inversion in the Colorado Plateau basin. *J. Appl. Meteor.*, **38**, 1103–1117, [https://doi.org/10.1175/1520-0450\(1999\)038<1103:WEOTTI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<1103:WEOTTI>2.0.CO;2).
- Wilson, T. B., and Coauthors, 2020: Evaluating time domain reflectometry and coaxial impedance sensors for soil observations by the U.S. climate reference network. *Vadose Zone J.*, **19**, e20013, <https://doi.org/10.1002/vzj2.20013>.
- Wulfmeyer, V., and Coauthors, 2018: A new research approach for observing and characterizing land–atmosphere feedback. *Bull. Amer. Meteor. Soc.*, **99**, 1639–1667, <https://doi.org/10.1175/BAMS-D-17-0009.1>.
- , J. M. V. Pineda, S. Otte, M. Karlbauer, M. V. Butz, T. R. Lee, and V. Rajtschan, 2023: Estimation of the surface fluxes for heat and momentum in unstable conditions with machine learning and similarity approaches for the LAFE data set. *Bound.-Layer Meteor.*, **186**, 337–371, <https://doi.org/10.1007/s10546-022-00761-2>.
- Xia, Y., M. B. Ek, Y. Wu, T. Ford, and S. M. Quiring, 2015: Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis. *J. Hydrometeorol.*, **16**, 1962–1980, <https://doi.org/10.1175/JHM-D-14-0096.1>.
- Xiu, A., and J. E. Pleim, 2001: Development of a land surface model. Part I: Application in a mesoscale meteorological model. *J. Appl. Meteor. Climatol.*, **40**, 192–209, [https://doi.org/10.1175/1520-0450\(2001\)040<0192:DOALSM>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0192:DOALSM>2.0.CO;2).